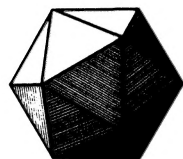


The American Mathematical Monthly



Volume 101, Number 1 / JANUARY 1994



Pierre de Fermat, 1601 – 1665

NOTICE TO AUTHORS

The *Monthly* publishes articles, notes, and other features about mathematics and the profession. The readership of the *Monthly* is intended to include everybody who is mathematically inclined, including of course professional mathematicians and students of mathematics at all collegiate levels. While no single article or feature is likely to appeal to everyone, material should interest and be accessible to a large number of readers. This is the most important criterion for acceptance.

Articles may be expositions of old results or presentations of new ones. They may concern all of mathematics or one small area, a broad development or a single application, historical reminiscences or one important event. While some articles may contain the author's new research, the novelty of material and generality of the results is far less important than the clarity of exposition and general interest. Discussing one illuminating case of a well known result is far better than providing all the details of an obscure but new proposition. Articles in the *Monthly* are supposed to inform and to entertain; they are meant to be read rather than archived.

Notes are short and possibly informal articles. A note may concern a clever new proof of an old theorem, a novel way to present tired material, or a lively discussion of a philosophical (but still mathematical) issue. Also, any topic is suitable, so long as it is related to mathematics. Because a note is short, the first few sentences are the most important part: They should explain the purpose and invite the reader in. Photographs or diagrams often will attract the reader's attention.

All articles and notes should be sent to the editor:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Please send 3 copies, typewritten on only one side of the paper. Illustrations should be carefully drawn on separate sheets of paper in black ink; the original should be without lettering and two copies should have appropriate captions and lettering indicated.

Proposed problems or solutions should be sent to:

RICHARD BUMBY,
P.O. Box 10971
New Brunswick, NJ 08906-0971.

Please send 2 copies of all material, typewritten if possible.

Letters to the Editor, both for publication and for private reading, should be sent to the Editor at the address given above. Comments, including criticisms, are welcome, as are all suggestions for making the *Monthly* a lively, entertaining, and informative journal.

EDITOR:

JOHN H. EWING

ASSOCIATE EDITORS:

RONALD BOOK	JOAN HUTCHINSON
PETER BORWEIN	FRED KOCHMAN
RICHARD BUMBY	CATHERINE MCGEOCH
DENNIS DETURCK	RICHARD NOWAKOWSKI
UNDERWOOD DUDLEY	ARNOLD OSTEBEE
JOHN DUNCAN	LEE RUBEL
JOAN FERRINI-MUNDY	LYNN STEEN
JOSEPH GALLIAN	STAN WAGON
STEVEN GALOVICH	DOUGLAS WEST
RICHARD GUY	HERBERT WILF
DARRELL HAILE	SANDY ZABELL
PAUL HALMOS	PAUL ZORN

EDITORIAL ASSISTANT:

MISTY CUMMINGS

STAFF ARTIST:

MIKE CAGLE

Reprint permission:

MARCIA P. SWARD, Executive Director

Advertising Correspondence:

Ms. ELAINE PEDREIRA, Advertising Manager

Subscription correspondence, change of address, and other inquiries:

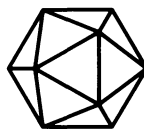
Membership / Subscriptions Department

All at the address:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036.

Microfilm Editions: University Microfilms International,
Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT. Copyrighted by the Mathematical Association of America (Incorporated), 1994, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source. Second class postage paid at Washington, DC, and additional mailing offices. **Postmaster:** Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.



Contents

ARTICLES

- Introduction to Fermat's Last Theorem / DAVID COX 3
Future Elementary Teachers: The Neglected Constituency /
THOMAS W. HUNGERFORD 15
Galois Theory for Beginners / JOHN STILLWELL 22
Into the Hourglass: Reflections on the Forces Acting on a Granular
Material / E. BRUCE PITMAN 28
Orderly Currencies / JOHN DEWEY JONES 36
An Interior Fixed Point Property of the Disc / ROBERT F. BROWN
and ROBERT E. GREENE 39
Le Cam's Inequality and Poisson Approximations /
J. MICHAEL STEELE 48
-

FEATURES

COMMENTS 2

PICTURE PUZZLE 27

NOTES

- A Generalization of a Theorem of Euler / DORINA MITREA
AND MARIUS MITREA 55
The Existence of a Triangle with Prescribed Angle Bisector Lengths /
PETRU MIRONESCU and LAURENTIU PANAITOPOL 58

THE COMPUTER SCIENCE SAMPLER

- Turing Machines and Computational Complexity / BILL MARION 61

THE EVOLUTION OF...

- The Evolution of Integration / A. SHENITZER and J. STEPRĀNS 66

THE AUTHORS 73

PROBLEMS AND SOLUTIONS 75

REVIEWS

- Mathematical Cranks*, by Underwood Dudley / IAN STEWART 87
Complex Analysis: The Geometric Viewpoint, by Steven G. Krantz /
JOHN POLKING 91

TELEGRAPHIC REVIEWS 95

COMMENTS

If you asked the average Athenian strolling across the Agora some fine day in 420 B.C., “How do you like living in the Golden Age?” the answer might have been, “What Golden Age? You think I don’t have troubles? And have you heard what that nut Pericles is up to now?” Residents in Golden Ages are not always aware of where they are. It may be that some readers of the *Monthly* are not aware that they are, right now, smack dab in the middle of the Golden Age of Mathematics. Evidence follows.

The first year of *Mathematical Reviews* contained 400 pages, index included. Five years later, after the backlog had been worked off, the length was 334 pages. One *month* (August) of *MR* in 1993 consumed 580 pages. I don’t think anyone will argue that the quality of papers reviewed has deteriorated sharply since 1940. Theorems are being proved now that have never been proved before. One owing to Fermat springs to mind.

In 1933, the Mathematical Association of America had, in the entire state of California, 84 members. The current *Combined Membership List* has 61 *columns* of members in California, with each column containing 90-odd names. Forty years ago, the MAA published one journal. Now there are three large ones and two or three smaller ones. Then, the Association published a few Carus Monographs and Slaughter Papers. Now, its catalog lists some 170 titles and is itself almost as long as a Slaughter Paper.

In 1990, 775,000 people were studying calculus in college. In 1993, 100,000 *high-school* students wrote the calculus advanced placement examination. It wasn’t too many years ago when there weren’t 100,000 people in the country who had ever had a course in calculus. Daniel Webster was not taught arithmetic in elementary school; now every member of Congress has been put through algebra. In 1990 there were more than 3,000,000 enrollments in college mathematics; allowing for duplication, that is almost 10% of the population! Mathematics and mathematicians command great prestige and respect, as is shown by how people will apologize for not being good at algebra. They don’t apologize for ignorance of metallurgy, even though metals figure as largely in the life of the average citizen as do *x*s and *y*s. I’m sure that future historians will look back in wonder at the craze for mathematics that swept the United States in the second half of the 20th century and speculate on its causes.

Of course, you hear grumblings about how students aren’t what they used to be. But they have *never* been what they used to be. In 1924, when the Leopold-Loeb case temporarily added the meaning “homosexual” to the word *moron*, a professor at the University of Chicago said he would need a new word for his students, and suggested *osseocaput*. Boneheads at the *University of Chicago*? In 1924? Ahmes the scribe probably complained in 1650 B.C. that Egyptian schools no longer turned out graduates able to deal with $2/17$ the way they used to. United States students seem always to finish in the top three of the Mathematical Olympiads. Students are all right. After all, we were students once, and look at us now.

In 1992, the National Science Foundation spent \$76.5 million for mathematical research. That’s \$2800 for each and every member of the American Mathematical Society. When you add in support for education and from other sources, every member of the AMS and the MAA got more than \$4000. If you belonged to both, you got more than \$8000. How much income tax did you pay last year?

We’ve got it all. The Golden Age is right now. We should realize it, appreciate it, and give thanks.

—Underwood Dudley

Introduction to Fermat's Last Theorem

David A. Cox*

The announcement last summer of a proof of Fermat's Last Theorem was an exciting event for the entire mathematics community. This article will discuss the mathematical history of Fermat's Last Theorem (which we will abbreviate throughout as FLT), broken up into the following periods:

1. Diophantus to Euler (250–1783 A.D.)
2. Euler to Frey (1783–1982 A.D.)
3. Frey to Wiles (1982–1993 A.D.)

We will give only an introduction to the story of Fermat's Last Theorem, and our account is by no means definitive. Many of the more technical terms are not defined completely and the few proofs that appear are only sketched. On the other hand, I hope that the article succeeds in conveying the flavor of this truly wonderful mathematics.

1. DIOPHANTUS TO EULER. Our history of FLT starts around 250 A.D. with Diophantus, whose *Arithmetica* considered many problems in elementary number theory. Consider Problem 8 from Book II, which asks “to divide a given square number into two squares” ([10], p. 144). Diophantus' solution is as follows: Let the given square be 16, let x^2 be one of the required squares and $(2x - 4)^2$ the other square. Therefore, we must satisfy $x^2 + (2x - 4)^2 = 16$, which implies

$$x^2 + 4x^2 - 16x + 16 = 16 \Rightarrow 5x^2 = 16x \Rightarrow x = 16/5.$$

Hence the required squares are $256/25$ and $144/25$.

We can observe two things about this problem. First, solutions are presumed to be rational. We neither restrict to only integer solutions nor generalize to real solutions. Second, we care only about finding one solution to a given problem; if we find one, we are happy and move on.

The *Arithmetica* was one of the last Greek mathematical works translated into Latin; this occurred in 1575. Fermat (1601–1665) had a copy of Bachet's translation of 1621 and made a series of intriguing annotations in its margins. Sometime in the late 1630's, while thinking about the problem given above, he added the famous words in the margin:

“On the other hand, it is impossible to separate a cube into two cubes, or a biquadrate into two biquadrates, or generally any power except a square into

*This article is based on a lecture given at the 1993 Smith College Regional Geometry Institute. This audience included high schools teachers, undergraduates, graduate students and researchers in discrete and computational geometry. I would like to thank Thomas Colthurst for transcribing the lecture, and I am grateful to my colleagues who pointed out errors in earlier versions of the manuscript.

two powers with the same exponent. I have discovered a truly marvellous proof of this, which however the margin is not large enough to contain.” ([10], pp. 144–145)

Hence the basic claim of Fermat’s Last Theorem is that the equation $x^n + y^n = z^n$ has no solutions when x, y, z are nonzero integers and $n > 2$. Generations of mathematical historians have debated over whether Fermat really did have a proof, though many experts doubt that he did. For one thing, the equation $x^n + y^n = z^n$ was atypical for Fermat—the vast majority of the other equations he studied dealt with exponents ≤ 4 . Also, in his correspondence, he only stated FLT for the exponent $n = 3$. As for Fermat’s “marvellous proof,” it probably used the technique of infinite descent. His descent argument for $n = 4$ is actually known: it can be found in Fermat’s proof that the area of a right triangle with integral sides cannot be a square. This proof is given in one of his marginal notes, although even here, Fermat complains that there isn’t enough room to give the proof “with all detail” ([10], p. 293). It seems likely that Fermat thought that his proofs for $n = 3$ and 4 generalized, and they almost certainly didn’t.

So, what happened after Fermat? In 1670, his marginal notes were published by his son, and some of his letters appeared in Wallis’ *Opera Mathematica*. In 1729, Goldbach wrote Euler and mentioned some of Fermat’s results. This got Euler, only 22 at the time, thinking about number theory. Three years later, Euler wrote his first paper on number theory, disproving a conjecture of Fermat’s on primes of the form $2^{2^n} + 1$. For the next fifty years, Euler proved many of Fermat’s conjectures and in so doing, transformed number theory from a collection of miscellaneous facts and results into an organized field at the very center of mathematics.

Here is an example of what Euler did. In Problem 17 of Book VI of the *Arithmetica* (Problem 19 in Bachet’s numbering), Fermat had written in the margin, “Can one find in whole numbers a square different from 25, when increased by 2, becomes a cube? . . . [The answer involves] the doctrine of whole numbers, which is assuredly very beautiful and very subtle . . .” ([6], p. 269). In modern terms, Fermat is claiming that the only integer solutions to $x^3 = y^2 + 2$ are given by $(x, y) = (3, \pm 5)$. You can see how the emphasis is different from Diophantus—Fermat is looking for *all* solutions, and he recognizes that asking for integer solutions (rather than rational ones) is a question of independent interest.

To prove Fermat’s claim, Euler ([4], Part II, §193) uses numbers of the form $a + b\sqrt{-2}$, with a, b integers. First observe

$$x^3 = y^2 + 2 = (y + \sqrt{-2})(y - \sqrt{-2}).$$

One can show that $y + \sqrt{-2}$ and $y - \sqrt{-2}$ are relatively prime, and since their product is a cube, each of them must also be a cube. Thus there is a number $p + q\sqrt{-2}$ such that

$$\begin{aligned} y + \sqrt{-2} &= (p + q\sqrt{-2})^3 = p^3 - 6pq^2 + (3p^3q - 2q^3)\sqrt{-2} \\ &\Rightarrow 1 = 3p^2q - 2q^3 = q(3p^2 - 2q^2). \end{aligned}$$

The last equation implies $p = \pm 1$ and $q = 1$. Substituting this in, we get $y = p^3 - 6pq^2 = \pm 5$ and $x = 3$, as claimed.

This proof, while elegant, is incomplete, for we do not know that numbers of the form $a + b\sqrt{-2}$ have *unique factorization*, or even for that matter, *primes* (although it is relatively easy to prove that the numbers $a + b\sqrt{-2}$ have these

properties). There are several reasons why this example is important:

- First, it reminds us that there are lots of diophantine equations besides just FLT, and what we really want is a method for dealing with as many of them as possible.
- Second, it shows that properties of integers (such as unique factorization) can apply in more general situations, and it illustrates how a result in one context (the integers) can be proved by working in a more general context (numbers of the form $a + b\sqrt{-2}$).
- Finally, the equation $y^2 = x^3 - 2$ is an example of an *elliptic curve*. Elliptic curves will play a crucial role in the final proof of FLT.

2. EULER TO FREY. This section is only a sketch of more than two hundred years of beautiful and wonderful number theory. For more information on the work on FLT done during this period, we warmly recommend both Edwards' *Fermat's Last Theorem* [3] and Ribenboim's *13 Lectures on Fermat's Last Theorem* [21]. (Precise references for results mentioned in this section can be found in these books.)

Before we begin, first observe that it suffices to prove FLT for $n = 4$ (done by Fermat) and for n an odd prime (since we can factor the exponent). We can also assume that x, y, z are nonzero relatively prime integers (because we can cancel common factors). That being said, here are some of the highlights of the 19th century work on FLT:

- By the early 1800s, all of Fermat's problems were solved except for FLT (thus justifying the name, Fermat's Last Theorem).
- 1816—The French Academy announces a prize for a solution to FLT.
- In the 1820's Sophie Germain shows that if p and $2p + 1$ are prime, then $x^p + y^p = z^p$ has no solution with $p \nmid xyz$. This is the so-called Case I of FLT. (Case II is where $p \mid xyz$ and is usually regarded as being much harder.)
- 1825—Dirichlet and Legendre prove FLT for $n = 5$.
- 1832—Dirichlet, after trying to prove it for $n = 7$, proves FLT for $n = 14$.
- 1839—Lamé proves FLT for $n = 7$.
- 1847—Lamé and Cauchy present false proofs of FLT for general n .
- 1844–1847—Kummer's work on FLT.

Let us describe Kummer's work on FLT in more detail. Kummer (and Cauchy and Lamé) started, à la Euler, by factoring the right hand side of the FLT equation as

$$x^p = z^p - y^p = (z - y)(z - \zeta y)(z - \zeta^2 y) \cdots (z - \zeta^{p-1} y),$$

where $\zeta = e^{2\pi i/p} = \cos(2\pi/p) + i \sin(2\pi/p)$ is a p th root of unity and satisfies $\zeta^p = 1$. In general, working with roots of unity will require us to use numbers of the form

$$a_0 + a_1 \zeta + \cdots + a_{p-1} \zeta^{p-1}, \quad a_0, \dots, a_{p-1} \in \mathbf{Z},$$

which are called *cyclotomic integers*. But a problem arises when unique factorization, one of our main tools, fails for the cyclotomic integers. As Kummer discovered in 1844, this first occurs for $p = 23$ (and in fact, unique factorization fails for all bigger primes as well).

Kummer's solution to this was twofold. First, he introduced a generalization of cyclotomic integers, called *ideal numbers*, which make up for the lack of unique

factorization. Second, he defined the *class number* h , which measures how badly unique factorization fails.

Here is a summary of Kummer's results:

- 1847—Theorem: FLT holds for p if $p \nmid h$ (such p are called *regular primes*).
- 1847—Theorem: p is regular iff p doesn't divide the numerator of the Bernoulli numbers B_2, B_4, \dots, B_{p-3} .

We can define the Bernoulli numbers by the power series

$$\frac{x}{e^x - 1} = \sum_{n=1}^{\infty} \frac{B_n}{n!} x^n.$$

A corollary of this result is that for $p < 100$, only 37, 59 and 67 are irregular.

- 1850—The French Academy offers a second prize for a solution to FLT.
- 1856—At Cauchy's suggestion, the Academy withdraws the prize and then awards a medal to Kummer.
- 1857—Kummer develops complicated criteria for proving FLT for certain irregular primes. There are some gaps in his proofs which are later filled in by Vandiver in the 1920s. These results establish FLT for $p < 100$.

The above history makes a wonderful story about how FLT inspired one of the greatest inventions in number theory, but the story is unfortunately false. Kummer was actually not trying to prove FLT, but something called a reciprocity theorem. Reciprocity theorems have their origins in Fermat's study of equations like $p = x^2 + y^2$ and $p = x^2 + 2y^2$, where p is a prime. In trying to understand these results, Euler, Lagrange, Legendre and Gauss created the theory of quadratic forms and proved the law of quadratic reciprocity. Later, Gauss, Abel and Jacobi formulated versions of cubic and biquadratic reciprocity, and Kummer and Eisenstein made the first attempts at higher reciprocity laws. Cyclotomic integers and ideal numbers came about primarily from Kummer's attempts to prove these higher reciprocity laws. In turn, these concepts not only had something interesting to say about FLT, but they also made significant contributions toward the development of class field theory and abstract algebra (we use the terminology "ideal of a ring" because of Kummer's "ideal numbers").

Here are some highlights of the history of FLT after Kummer:

- 1908—The Wolfskehl prize for a solution to FLT is announced. Later inflation in the German mark reduces the value of this prize considerably, but does not reduce the flow of crank solutions submitted.
- 1909—Wieferich proves if $x^p + y^p = z^p$ and $p \nmid xyz$ (Case I of FLT), then $2^{p-1} \equiv 1 \pmod{p^2}$. This is a strong congruence which is particularly easy to check on a computer.
- 1953—Inkeri proves that if $x^p + y^p = z^p$ and $x < y < z$, then $x > ((2p^3 + p)/\log(3p))^p$ in Case I and $x > p^{3p-4}$ in Case II.
- 1971—Brillhart, Tonascia and Weinberger show that Case I of FLT is true for all primes less than $3 \cdot 10^9$.
- 1976—Wagstaff shows that FLT is true for all primes less than 125,000.

These results imply that any counterexample to FLT must involve $p \geq 125,003$ and $z > y > x > (125,003)^{375,005} \approx 4.5 \cdot 10^{1,911,370}$. (In 1992, as a byproduct of other computations, the lower bound on the exponent was raised to $p > 4,000,000$ —see [2].)

We should also mention that the Fermat equation $x^n + y^n = z^n$ has been studied in many other contexts, including polynomials, entire functions and matrices (see [21] and, for a recent proof of the polynomial case, [18]).

3. FREY TO WILES. In 1983, Faltings [4] proved the Mordell Conjecture, which implies that a polynomial equation with rational coefficients $Q(x, y) = 0$ has only finitely many rational solutions when the curve has genus ≥ 2 (for a definition of genus, see the sidebar “The genus of an algebraic curve”). Since $x^n + y^n = 1$ has genus ≥ 2 for $n \geq 4$, there are only finitely many rational solutions by the Mordell Conjecture. Then, clearing denominators, it follows easily that $x^n + y^n = z^n$ has only finitely many relatively prime integer solutions.

The genus of an algebraic curve

The *genus* of a curve given by a polynomial equation $p(x, y) = 0$ of degree n can be defined in a variety of ways. When the equation is sufficiently smooth (which is true for the Fermat curve $x^n + y^n = 1$), then the genus is $g = (n - 1)(n - 2)/2$. This is ≥ 2 when $n \geq 4$.

Topologically, the solutions of $p(x, y) = 0$ over the complex numbers form a compact Riemann surface minus a finite set of points, and then the genus is just the usual genus of this compact real 2-dimensional manifold.

Analytically, a Riemann surface is a compact complex 1-dimensional manifold, and one can define the notion of a *holomorphic 1-form*. Then the genus is the maximum number of linearly independent holomorphic 1-forms on the surface.

For example, the Riemann sphere has genus zero, so that there are no holomorphic 1-forms, while the elliptic curve $y^2 = Ax^3 + Bx^2 + Cx + D$ has genus 1, and up to a constant, dx/y is the only holomorphic 1-form.

This may not seem so useful, since we want to show that the number of solutions is actually zero. But Granville [9] and Heath-Brown [11], aided by an observation of Filaseta, used the above finiteness result to show that FLT holds for “most” exponents, in the sense that if you look at all exponents—prime and composite—from 3 to n , the percentage where FLT could fail approaches zero as n increases (see [28] for the details). Also, Adelman and Heath-Brown [1] showed that Case I of FLT was true for infinitely many prime exponents.

For us, the Mordell Conjecture is interesting because it shows how a general conjecture in number theory can have some consequences concerning FLT. Also, in proving the Mordell Conjecture, Faltings used the machinery of modern algebraic geometry, which had been developing since the 1950’s.

By the end of the 1980’s, there were several conjectures in number theory which, if proved, would imply FLT, though sometimes only for sufficiently large exponents (see the sidebar “Conjectures that imply Fermat’s Last Theorem”). This showed that FLT was not an isolated oddity, but rather was intimately connected to other parts of number theory. People were especially excited in 1988, when Miyaoka gave a lecture in Bonn in which he stated one of these conjectures, the arithmetic Bogomolov-Miyaoka-Yau inequality, as a theorem. This would have proved FLT for all large primes p (without saying explicitly what “large” meant). In the days following his lecture, there was much fanfare in the press, and it was rather disappointing when a week later an error was found in the argument.

Conjectures that imply Fermat's Last Theorem

By the late 1980's, there were several conjectures in number theory which, if proved, would imply FLT, at least for large exponents:

Diophantine Geometry. The following conjectures in diophantine geometry would imply FLT for all sufficiently large exponents:

- The *abc* Conjecture states that if a, b, c are relatively prime integers with $a + b = c$, then $\max(|a|, |b|, |c|)$ is bounded in terms of the primes dividing abc .
- Szpiro's Conjecture relates the minimal discriminant to the conductor of an elliptic curve. These terms are discussed in the sidebar entitled "Invariants of the Frey curve".
- Vojta's Conjecture concerns heights of points (relative to the canonical class) of a curve defined over the integers.

Precise statements of these conjectures (and the relations among them) can be found in Lang's survey article [18].

Arithmetic Surfaces. The Bogomolov-Miyaoka-Yau inequality for arithmetic surfaces relates various invariants of a curve defined over the integers. This inequality is an arithmetic analog of a well known inequality for complex surfaces. By Parshin [20] and Vojta (Appendix to [16]), this conjecture implies versions of the above diophantine conjectures strong enough to prove FLT for all large exponents.

Elliptic Curves. The Taniyama-Shimura Conjecture states that all elliptic curves over the rational numbers are modular (a more precise statement of the conjecture is in the body of the article). As we will explain, the work of Frey, Serre and Ribet shows that this conjecture implies FLT for *all* exponents.

Of these conjectures, the one ultimately most important for FLT is the Taniyama-Shimura Conjecture, which asserts that all elliptic curves over \mathbf{Q} are modular (this term will be defined below). The full story of this conjecture goes back to Jacobi and Riemann, but we will begin our account with the work of Gerhard Frey from 1982 to 1986 (see [7] and [8]). Frey showed that nontrivial solutions to FLT give rise to very special elliptic curves, which we shall call *Frey curves*. His basic insight was that Frey curves were so special that they couldn't be modular. Hence, if the Taniyama-Shimura Conjecture were true, Frey curves couldn't exist, and FLT would follow.

If $a^p + b^p = c^p$ is a solution to FLT, then the associated Frey curve is

$$y^2 = x(x - a^p)(x + b^p).$$

As usual, we assume a, b, c are nonzero relatively prime integers and p is an odd prime. This is an elliptic curve over the rational numbers \mathbf{Q} , similar to the equation $y^2 = x^3 - 2$ considered by Fermat. In general, an *elliptic curve over \mathbf{Q}* is given by an equation of the form

$$y^2 = Ax^3 + Bx^2 + Cx + D,$$

where A, B, C, D are rational and the cubic polynomial in x on the right hand side of the equation has distinct roots. Elliptic curves are a large and important part of modern number theory.

Actually, we have to be a bit careful when constructing the Frey curve. A solution $a^p + b^p = c^p$ gives rise to solutions $b^p + a^p = c^p$ and $a^p + (-c)^p = (-b)^p$ (since p is odd). From here it is easy to rearrange the solution so that b is even and $a \equiv -1 \pmod{4}$. This is needed in order that the Frey curve be *semistable* (this concept will be discussed below). For technical reasons, we will also assume that $p > 3$.

Although Frey had published a paper about Frey curves in 1982 [8], things didn't get really interesting until 1985, when Frey tried to prove that the Taniyama-Shimura Conjecture implies FLT. But his proof had some serious gaps. Several people tried to fix Frey's argument, and it was Jean-Pierre Serre [24] who saw that a special version of a conjecture he made on level reduction for modular Galois representations would fill the gap. Hence we may credit Frey and Serre with showing that FLT follows from Taniyama-Shimura and the special level reduction conjecture made by Serre. Versions of this argument can be found in [7] and [22] and one should also consult Serre's article [25].

Then, in 1986, Ken Ribet made significant progress along this route to FLT by proving this version of Serre's conjecture, and his proof eventually appeared in [22]. Thus FLT (for *all* primes p) was now a consequence of the Taniyama-Shimura Conjecture! Inspired by this development, Andrew Wiles began to work on Taniyama-Shimura, and seven years later, he presented a proof on June 23, 1993 that the conjecture is true for semistable elliptic curves, which (as we will see below) is good enough to prove FLT. Wiles' argument is not easy—the manuscript containing the proof is over 200 pages long. But many people in the mathematical community are confident that the proof will hold up under careful scrutiny. For a broad outline of Wiles' argument, see Ribet's article [23] (this article also has some useful references).

One interesting observation is that Frey was not the first to discover the Frey curve. On page 262 of [12], Hellegouarch writes down the Frey curve for a solution to FLT of exponent $2p^h$. Frey curves also appear implicitly as part of the correspondence between Fermat curves and modular curves considered by Kubert and Lang [15]. But Frey was clearly the first to suspect that the Frey curve couldn't exist because of the Taniyama-Shimura Conjecture.

To explain the Taniyama-Shimura Conjecture, we first need to define the concept of modular function.

Definition. A function $f(z)$ on the upper half plane $\{z = x + iy: y > 0\}$ is a modular function of level N if $f(z)$ is meromorphic, even at the cusps (see the sidebar “The modular curve $X_0(N)$ ”), and for all integers a, b, c, d with $ad - bc = 1$ and $N|c$, we have

$$f\left(\frac{az + b}{cz + d}\right) = f(z).$$

Taniyama-Shimura Conjecture. Given an elliptic curve $y^2 = Ax^3 + Bx^2 + Cx + D$ over \mathbf{Q} , there are nonconstant modular functions $f(z), g(z)$ of the same level N such that

$$f(z)^2 = Ag(z)^3 + Bg(z)^2 + Cg(z) + D.$$

Thus the Taniyama-Shimura Conjecture says that an elliptic curve over \mathbf{Q} can be parameterized by modular functions, or, as Mazur says in [19], it has a “hyperbolic uniformization.” Such an elliptic curve is said to be *modular*. Wiles

The modular curve $X_0(N)$

In the text, we considered the transformations $z \mapsto (az + b)/(cz + d)$ of the upper half plane $\mathfrak{h} = \{z = x + iy: y > 0\}$ associated to the group of matrices

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbf{Z}, ad - bc = 1, N|c \right\}.$$

This group acts on \mathfrak{h} with quotient $\mathfrak{h}/\Gamma_0(N)$, and there is a compact Riemann surface $X_0(N)$ such that

$$\mathfrak{h}/\Gamma_0(N) = X_0(N) - \{\text{finite set of points}\}.$$

These points are the *cusps* and $X_0(N)$ is a *modular curve of level N* . (Other modular curves come from using different groups of matrices).

A function $f(z)$ is invariant under $\Gamma_0(N)$ iff it descends to a function on $\mathfrak{h}/\Gamma_0(N)$. Thus the definition of modular function means that we have a meromorphic function on $X_0(N)$. Furthermore, the Taniyama-Shimura Conjecture asserts that if E is an elliptic curve over \mathbf{Q} , then there is a surjective holomorphic map $X_0(N) \rightarrow E$.

As suggested in the text, a holomorphic 1-form on $X_0(N)$ is written $F(z) dz$, where $F(z)$ is a cusp form of weight 2 and level N . It follows that the genus of $X_0(N)$ (see the sidebar “The genus of an algebraic curve”) equals the dimension of the space of these cusp forms.

It is known that $X_0(2)$ has genus zero (this follows by looking at the fundamental domain of $\Gamma_0(2)$ acting on \mathfrak{h}). Hence *there are no cusp forms of weight 2 and level 2*. This fact is used in the proof of FLT.

The modular curves $X_0(N)$ play an important role in the theory of elliptic curves. Some basic facts about modular curves can be found in [14] and [26] (and other references can be found in these books).

proved this conjecture for semistable elliptic curves. We should mention that our statement of the conjecture is very naive—some work is needed to show it is equivalent to the usual formulation (see the technical appendix to [19]). For a discussion of the conjecture and some of its history, see pages 130–135 of Lang’s book [17]. At a more elementary level, Mazur’s article “Number theory as gadfly” [19] gives a lovely introduction to the Taniyama-Shimura Conjecture.

Besides modular functions, we also need to know about modular forms of weight 2. The easiest way to see how these arise is through elliptic integrals. An elliptic integral is an integral of the form

$$\int \frac{dx}{\sqrt{Ax^3 + Bx^2 + Cx + D}}.$$

(Strictly speaking, this is only an elliptic integral of the first kind—there are many other types of elliptic integrals.) If $y^2 = Ax^3 + Bx^2 + Cx + D$, then this integral is simply $\int dx/y$. For a modular elliptic curve, we have $x = f(z)$, $y = g(z)$, and then

$$\frac{dx}{y} = \frac{df}{g} = \frac{f'(z) dz}{g(z)} = F(z) dz.$$

One can show that for a, b, c, d as in the definition of modular function, the above function $F(z)$ transforms via the rule

$$F\left(\frac{az + b}{cz + d}\right) = (cz + d)^2 F(z).$$

We call $F(z)$ a *modular form of weight 2 and level N* . When the modular parametrization is chosen correctly, the function $F(z)$ has some remarkable properties. It is holomorphic and vanishes at the cusps, and for this reason is called a *cusp form*. In addition, $F(z)$ is an *eigen-form* for the action of a certain Hecke algebra on the space of all cusp forms. So $F(z)$ is a rather sophisticated object.

The miracle is that $F(z)$ is intimately connected to the curve $y^2 = Ax^3 + Bx^2 + Cx + D$. Roughly speaking, one can reconstruct $F(z)$ simply by knowing the number of solutions of the congruences $y^2 \equiv Ax^3 + Bx^2 + Cx + D \pmod{p}$ for all primes p . Then the fact that $F(z)$ is a cusp form of weight 2 and level N tells us some profound things about the elliptic curve. This is one reason why Taniyama-Shimura is such a wonderful conjecture—number theorists would be excited by its proof even if there were no connection to FLT.

We can now sketch the argument of Frey and Serre which shows why FLT follows from Taniyama-Shimura and the level reduction conjecture of Serre. We begin with the FLT solution $a^p + b^p = c^p$. As above, we will assume $p > 3$ is prime and a, b, c are relatively prime with b even and $a \equiv -1 \pmod{4}$. We then get the Frey curve $y^2 = x(x - a^p)(x + b^p)$.

The *discriminant* of a polynomial is the product of the squares of the differences of its roots. For the cubic $x(x - a^p)(x + b^p)$, the discriminant equals

$$(a^p - 0)^2(-b^p - 0)^2(a^p - (-b^p))^2 = a^{2p}b^{2p}c^{2p}$$

Invariants of the Frey curve

Suppose that we have the Frey curve $y^2 = x(x - a^p)(x + b^p)$ with our usual assumptions on a, b, c . Then we get the following invariants:

- Besides the discriminant defined in the text, an elliptic curve over \mathbf{Q} has a more subtle invariant called the *minimal discriminant*. The minimal discriminant of the Frey curve is $\Delta = 2^{-8}a^{2p}b^{2p}c^{2p}$. Since b is even and $p \geq 5$, this is still an integer. This differs from the discriminant because the discriminant depends on the particular equation defining the curve, while the minimal discriminant is intrinsic to the curve.
- The *conductor* of the Frey curve is $N = \prod_{l|abc} l$. The conductor is the most subtle of the invariants associated to an elliptic curve over \mathbf{Q} . One can show that a modular elliptic curve is parametrized by modular functions whose level N equals the conductor of the curve.
- The *j -invariant* of the Frey curve is $j = 2^8(a^{2p} + b^{2p} + a^pb^p)^3/a^{2p}b^{2p}c^{2p}$. The j -invariant classifies the curve up to isomorphism over the complex numbers.

Precise definitions of these invariants can be found in Silverman's book [26] (see pp. 48, 224 and 361). The calculations for the Frey curve can be found in [7] and [25].

since a, b, c is a solution to FLT. It is unusual for a discriminant to be a pure $2p^{\text{th}}$ power—this is our first hint that the Frey curve is very special.

Besides the discriminant of its equation, an elliptic curve over \mathbf{Q} has a variety of invariants, including its *minimal discriminant* Δ , *conductor* N and *j-invariant* j . For the Frey curve, these invariants are given in the sidebar “Invariants of the Frey curve.” In general, Δ , N and j give useful information about the elliptic curve. For instance, when the curve is modular, one can find a modular parametrization using modular functions of level N , where N is the conductor of the curve. This fact will play an important role in the proof below.

We then have the following results about the Frey curve:

Lemma 1. *The Frey curve is semistable.*

Proof: We first need to define what semistable means. When a prime l divides the discriminant, two or possibly all three of the roots become congruent modulo l . Roughly speaking, an elliptic curve is semistable if for all such primes l , only two roots become congruent mod l (the definition is more complicated for the primes 2 and 3). Thus, for primes bigger than 3, the Frey curve is semistable since the discriminant is $a^{2p}b^{2p}c^{2p}$ and the roots are 0, a^p and $-b^p$, where a^p and b^p are relatively prime. More work is required to check semistability at $l = 2$ or 3, and when $l = 2$, the conditions b even, $a \equiv -1 \pmod{4}$ and $p > 3$ are needed. For more details, see [7] and [25]. Q.E.D.

Corollary (Wiles). *The Frey curve is modular.*

Lemma 2. *For every odd prime l dividing N , the j -invariant of the Frey curve can be written as $j = l^{-mp} \cdot q$, where m is a positive integer and q is a fraction not involving l . (We say that the j -invariant is exactly divisible by l^{-mp} in this case.)*

Proof: The j -invariant of the Frey curve is

$$\frac{2^8(a^{2p} + b^{2p} + a^p b^p)^3}{a^{2p} b^{2p} c^{2p}} = \frac{2^8(c^{2p} - b^p c^p)^3}{(abc)^{2p}}.$$

The power of l dividing the denominator is obviously a multiple of p , and since a, b, c are relatively prime, one sees that $(c^{2p} - b^p c^p)^3$ and $(abc)^{2p}$ are relatively prime. Since N is the product of the primes dividing abc , the lemma follows easily. The lemma fails for $l = 2$ because of the factor of 2^8 in numerator. Q.E.D.

In the context of these three results—semistable modular elliptic curves whose j -invariants are exactly divisible by $l^{-\text{multiple of } p}$ for odd primes l dividing N —the level reduction conjecture of Serre applies for *all* odd primes dividing N (see [6] and [25] for the details of how this works). Serre’s conjecture involves Galois representations and is rather technical (see [22] for a precise statement), though we will discuss its implications below.

We can now prove Fermat’s Last Theorem:

Theorem. *The equation $x^p + y^p = z^p$ has no solutions with a, b, c nonzero for p an odd prime.*

5. G. Faltings, Endlichkeitssätze für abelschen Varietäten über Zahlkörpern, *Invent. Math.* 73 (1983), 349–366. For an English translation, see *Arithmetic Geometry*, edited by G. Cornell and J. Silverman, Springer-Verlag, Berlin Heidelberg New York, 1986, 9–27.
6. de Fermat, Pierre, *Oeuvres de Fermat*, Volume 3, Gauthier-Villars, Paris, 1896.
7. G. Frey, Links between stable elliptic curves and certain diophantine equations, *Ann. Univ. Sarav.* 1 (1986), 1–40.
8. G. Frey, Rationale Punkte auf Fermatkurven und getwisteten Modulkurven, *J. reine u. angew. Math.* 331 (1982), 185–191.
9. A. Granville, The set of exponents, for which Fermat’s Last Theorem is true, has density one, *Comptes Rendus/Mathematical Reports*, Academy of Science, Canada, 7 (1985), 55–60.
10. T. Heath, *Diophantus of Alexandria*, Second Edition, Cambridge University Press, Cambridge, 1910. (Reprint by Dover Books, New York, 1964.)
11. D. Heath-Brown, Fermat’s Last Theorem is true for “almost all” exponents, *Bull. Lon. Math. Soc.* 17 (1985), 15–16.
12. Y. Helloguarch, Points d’ordre $2p^h$ sur les courbes elliptiques, *Acta Arith.* 26 (1975), 253–263.
13. N. Koblitz, *Introduction of Elliptic Curves and Modular Forms*, Springer-Verlag, Berlin Heidelberg New York, 1984.
14. A. Knapp, *Elliptic Curves*, Princeton University Press, Princeton, 1992.
15. D. Kubert and S. Lang, Units in the modular function field, I, *Math. Ann.* 218 (1975), 67–96.
16. S. Lang, *Introduction to Arakelov Theory*, Springer-Verlag, Berlin Heidelberg New York, 1988.
17. S. Lang, *Number Theory III: Diophantine Geometry*, Springer-Verlag, Berlin Heidelberg New York, 1991.
18. S. Lang, Old and new conjectured diophantine inequalities, *Bull. AMS* 23 (1990), 37–75.
19. B. Mazur, Number theory as gadfly, *Am. Math. Monthly* 98, 593–610.
20. A. Parshin, The Bogomolov-Miyaoka-Yau inequality for the arithmetical surfaces and its applications, *Séminaire de Théorie des Nombres*, Paris, 1986–87.
21. P. Ribenboim, *13 Lectures on Fermat’s Last Theorem*, Springer-Verlag, Berlin Heidelberg New York, 1979.
22. K. Ribet, On modular representations of $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ arising from modular forms, *Invent. Math.* 100 (1990), 431–476.
23. K. Ribet, Wiles proves Taniyama’s Conjecture; Fermat’s Last Theorem follows, *Notices AMS* 40 (1993), 575–576.
24. J.-P. Serre, Lettre à J.-F. Mestre, in *Current Trends in Arithmetical Algebraic Geometry*, Contemporary Mathematics 67, AMS, Providence, 1987, 263–268.
25. J.-P. Serre, Sur les représentations modulaires de degré 2 de $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, *Duke Math. J.* 54 (1987), 179–230.
26. J. Silverman, *The Arithmetic of Elliptic Curves*, Springer-Verlag, Berlin Heidelberg New York, 1986.
27. J. Silverman and J. Tate, *Rational Points on Elliptic Curves*, Springer-Verlag, Berlin Heidelberg New York, 1992.
28. S. Wagon, The evidence: Fermat’s Last Theorem, *Math. Intelligencer* 8, No. 1 (1986), 59–61.

Department of Mathematics & Computer Science
Amherst College
Amherst, MA 01002
dac@cs.amherst.edu

Who Was the Author?

Zur Invariantentheorie der Formen von n Variablen, J. Ber. d. DMV, 1910.

Idealtheorie in Ringbereichen, Math. Ann., 1921.

Hilberstsche Anzahlen in der Idealtheorie, J. Ber. d. DMV, 1925.

Nichtkommutative Algebren, Math. Z. 1933.

Answer on page 54

Future Elementary Teachers: The Neglected Constituency*

Thomas W. Hungerford

“We have met the enemy and he is us.”

—Pogo Possum

WHERE WE ARE NOW. Complaints about the mathematical preparation of incoming students are endemic in both college and high school mathematics departments. College professors tend to blame the high school teachers and they, in turn, blame the elementary school teachers for the sad state of affairs. Although this linear model of blame may be comforting to those at the top, a circular model (with blame for all) is a much better reflection of reality because elementary teachers are “trained” by the same college professors who complain about incoming students. In the words of J. R. C. Leitzel [3],

The mathematical preparation of elementary school teachers is perhaps the weakest link in our nation’s entire system of mathematics education.

Colleges and universities have done little to improve the situation because neither the students (prospective elementary teachers) nor their instructors have any immediate reason to change. Many of the students have weak mathematical backgrounds and a high level of mathematical anxiety. The last thing they want is to take more mathematics courses or “harder” ones.

Many instructors who teach courses for elementary teachers do so unwillingly (I avoided teaching them for almost twenty years). If they can’t teach an advanced course, they would rather teach calculus or even precalculus—courses with some modicum of “real mathematics.” It’s no wonder they feel this way. With few exceptions, the typical “mathematics for elementary teachers” sequence is taught from one of several isomorphic textbooks of the same title and purports to “cover” all of the mathematics in grades K–8 in approximately six semester hours. Every mathematical term that might possibly appear in elementary school is mentioned, with depth of coverage being inversely proportional to this breadth.

Apparently, many instructors take the path of least resistance and follow the textbook, bad as it may be. From one point of view, this is not particularly

* This article arises from my experiences at a major research university and a comprehensive urban university. In both departments, the mathematics courses for elementary teachers are routinely taught by mathematicians whose primary interests are not in mathematics education. Nevertheless, I suspect that the comments here apply equally well (with occasional minor changes) to all but a handful of the colleges and universities in this country that offer special courses for prospective elementary teachers.

surprising. Innovative teaching requires time and time given to teaching is time taken from research. With the reward system being what it is in many departments, there is little incentive for faculty members to attempt improvements.

This cannot be the sole explanation, however. Despite similar pressures, calculus reform is thriving and a great deal is being done to change the teaching of linear algebra, differential equations and a variety of other courses. Nor is reform restricted to courses taken by math majors (in which all of us, understandably, have a vested interest). For instance, at the 1992 Joint Mathematics Meetings in Baltimore there were two sessions devoted to innovations in mathematics courses for business students and three sessions to liberal arts mathematics courses.

Considering what is being done to improve other courses, the lack of attention being paid to mathematics courses for prospective elementary teachers is astounding. Of more than 150 “math education” papers presented at the Baltimore meetings, only one specifically mentioned elementary teachers in its title. Anecdotal evidence suggests that this proportion accurately reflects the larger mathematical community’s interest in future elementary teachers as well.

The operative word here is “future.” There are a number of mathematics programs for in-service elementary teachers run by various universities and urban mathematics collaboratives, with funding from NSF, Exxon, and other foundations. Similar innovative projects for future teachers are extremely rare. Programs for in-service teachers are absolutely essential and will continue to be so for some time, but they have no direct effect on future teachers—those now enrolled in “math for elementary teachers.” This investment in the present generation of teachers will be wasted if we continue to ignore the generation that will eventually replace them.

WHERE WE ARE HEADED. Meeting society’s increasing need for a mathematically literate work force clearly requires more students to choose careers in mathematics and science. The professional well-being of those who teach mathematics depends in significant part on having a sufficient number of math majors to enable our departments to offer a variety of interesting courses and to obtain adequate support from the administration. Building and sustaining a political climate sympathetic to mathematical research, which will encourage government to provide more funding for researchers, depends on the public’s having at least a minimal appreciation of what mathematics is and what mathematicians do.

There is a growing body of evidence to suggest that students’ attitudes toward mathematics and science are well established by the time they *enter* high school. If they haven’t developed positive attitudes by then, the chances of their considering a career (or college major) in math or science is practically nil, as is the likelihood of their later supporting increased funding for mathematical research. Obviously, the elementary teacher plays a key role in developing a student’s appreciation of mathematics.

Elementary teachers who don’t know much mathematics, who have little interest in what it means to do mathematics, and who are afraid of mathematics, are not likely to engender positive attitudes toward mathematics in their students. Yet these are the kinds of teachers that the current system is geared to produce. The inescapable corollary is that *continued neglect of the mathematical education of prospective elementary school teachers courts disaster, both for the mathematical profession and the larger society.*

WHERE WE OUGHT TO GO INSTEAD. Making the changes needed to turn things around will require both curricular reform and attitude adjustment. As to the first, there is no secret about what ought to be done. A general framework is given in the M.A.A. report *A Call For Change: Recommendations For The Mathematical Preparation Of Teachers Of Mathematics* [4], the first half of which deals with elementary teachers, and in pages 132–139 of *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics) [5]. Contrary to much current practice, the recommendations assume that mathematics departments must

- take seriously the needs of future elementary teachers;
- offer them courses that are adequate in terms of both breadth and depth; and
- expect them to achieve at a reasonable level.

Both *A Call for Change* and the relevant pages of the NCTM *Standards* can easily be read at one sitting, so there is no need to paraphrase them here. It may be helpful, however, to consider what implementing these recommendations will require.

Level. The current prerequisites for “Math for Elementary Teachers” are usually the absolute minimum allowable (two years of college preparatory high school math is typical). The course work recommended in *A Call For Change* presumes three years of college preparatory high school mathematics for prospective teachers in grades K–4 and four years for those in grades 5–8. Consequently, students without this preparation (most of them at present) may have to take additional courses before they begin the teachers’ sequence.

Length. At most colleges and universities prospective elementary teachers are currently required to take approximately 6 semester hours (or less) of mathematics. The recommended minimums are 9 semester hours for teachers in grades K–4 and 15 semester hours in grades 5–8.

Content. In terms of breadth, the recommendations include virtually all of the major topics mentioned in current texts, plus some that are not (such as algebraic structures and concepts of calculus). However, since 50% to 150% more time is to be devoted to this material than at present, the coverage will have considerably more depth.

Flavor. With few exceptions, the additional credit hours will require not simply more courses from those already offered for other audiences, but some new courses specifically designed for this audience. In addition to mathematical content, these students should learn what it means to “do mathematics”—that doing mathematics is much more than just getting the right answers to computational problems.

Pedagogy. In common with almost all current reform efforts in collegiate mathematics, the recommendations call for a change in *how* mathematics is taught to prospective teachers. It won’t be enough simply to design new courses or select new texts. Instructors in these (and other) courses must be encouraged to adopt teaching styles other than the conventional lecture. Ideally, the passive model used in most current courses (I tell you what mathematics you need) will be replaced by

an active one (with my assistance, you explore, discover, or construct the mathematics you need).

Technology. Future teachers will need to be familiar and comfortable with a variety of technological tools, including ordinary calculators, graphing calculators, and computers, as well as appropriate geometric and computational software. It will take time, effort, and money for mathematics departments to develop and/or purchase suitable materials and to incorporate them into mathematics courses.

WHAT IT WILL TAKE TO GET THERE. Obviously, full implementation of the M.A.A. recommendations will take considerable time and work. Those who do it may have to take on not only their departmental colleagues, but also the college of education, the university bureaucracy, and possibly the state department of education. Such restructuring is unlikely to happen, however, unless there is a significant change in attitude on the part of individuals, departments, and institutions.

To begin with, mathematicians will have to abandon the belief that mathematics education in general, and certainly anything involving elementary teachers, is not their business. Such an attitude is reminiscent of the weapons scientist in the Tom Lehrer song (“once the rockets are up, who cares where they come down; that’s not my department . . .”) and equally absurd. No one else has as great a personal stake in the mathematical education of future elementary teachers as we do. If we don’t make their education our business, the rockets will come down on us.

It isn’t necessary, or even desirable, for every mathematician to be involved in reforming the mathematics program for future teachers. But it is essential that those who have the intelligence, interest and talent to do this well be encouraged to do so. H. Clemens [1] has addressed these issues in a broader context, but his remarks are equally applicable here:

This will demand . . . that research mathematicians show more respect for the enormous difficulties of mathematics education in the setting of a mass nonelite audience, and more respect for the intelligence and precious talent of educators who manage to achieve success in that setting.

Such respect must be more than just lip service; it must include rewards, in terms of salary, promotion, teaching loads, etc., comparable to those given good researchers. This does not mean that *any* work with elementary teachers is worthy of reward—mediocrity there is no more deserving than mediocrity in research. But it does mean that people doing first-rate work in *either* area ought to be adequately rewarded. Encouraging appropriate individuals to undertake the necessary work and rewarding those who do it well are primarily departmental responsibilities. Carrying out these responsibilities successfully will require leadership from the chairperson and senior members of the department, and possibly a significant change in the departmental ethos.

A new attitude in the department is necessary, but far from sufficient. Effective change won’t be possible unless the education college requires its elementary majors to participate in the reformed program. Such action cannot be taken for granted. Many education faculty members do not value mathematics to the same degree that we do and are satisfied with the present situation. If mathematicians must be convinced of the real need for change, we can hardly expect education

faculty to be any different. Most of us are unaccustomed to having to “sell” mathematics to others, but in this case it is essential.

Concerns about academic turf may further complicate the reform process. Within a university, requiring more courses in one area may generate pressure to require fewer in other areas. At many institutions, the amount of mathematics currently required of elementary education majors is a reflection of the minimum state certification requirements for elementary teachers. Because of real or perceived fears of losing students to competitive institutions, the certification requirements may be used as an argument against increasing the amount of mathematics required for future teachers; (if we do more than the state requires, students will go to schools that don’t). Depending on where you live, this could be a very serious problem. Of 54 states and territories* that certify teachers,

More than one-fourth require no mathematics at all for elementary certification!

Furthermore, as best I can determine,

*At most 10% require more than 6 semester hours of mathematics for elementary certification.***

These figures are depressing enough when “elementary” is defined as K–5 or K–6, but in many states elementary certification allows a person to teach seventh and eighth grade as well. Thus, revitalizing the mathematical education of future elementary teachers may require not only the consent of the mathematics department and the cooperation of the college of education, but also institutional support at the highest level.

WHAT TO DO UNTIL THE DOCTOR COMES. It would be unrealistic to expect large numbers of mathematicians to make reform of the program for elementary teachers their first priority. Nevertheless, in all but the smallest departments, there are likely to be some capable people who would be willing to do that. Those not directly involved can certainly encourage and support their efforts.

If you have never taught the course, you might consider doing so. Teaching the course even once will give you a much better picture of what the problems are and why change is necessary. In any case, it will take time to bring about change. In the meantime, there are many small things that any instructor can do to improve the present course. Here, in no particular order, are a few possibilities based on my own experience.

Make your students aware of the mathematical resources available to them, particularly journal articles at their level. I routinely require my students to write an “article report” on something they have read in *The Arithmetic Teacher* or other expository journals. Until they go looking for such an article, they are unaware of

*Fifty states, the District of Columbia, Puerto Rico, Guam, and the Virgin Islands.

**Ten states do not specify minimum numbers of credit hours, but approve an institution’s entire program for teachers. In Ohio (one of the ten), the vast majority of institutions require no more than 6 semester hours of mathematics. It seems likely that the same is true of the other nine.

the vast amount of mathematical literature designed for classroom teachers and of the practical help such articles can provide.

If possible, have your students attend a meeting of the state or local affiliate of the National Council of Teachers of Mathematics. Recently, the annual meeting of the Ohio Council of Teachers of Mathematics was held here, so I required my students to register (\$5) and attend at least one session. Most of them were amazed to discover how much is going on in elementary school mathematics. They were also impressed with the exhibits by publishers and school-supply organizations—resources most of them had been unaware of.

Try to wean yourself from the traditional lecture method, at least for part of each class. I have found that a relatively painless way to do this is to have the students spend the first 10–15 minutes of class in groups of 3 or 4 discussing current and past homework while I circulate from group to group, giving suggestions or clarifying what needs to be done. I don't work the problems for them, but encourage them to "talk about it with the group." At first some students are unhappy that I won't just give them the answer, but they soon find that the vast majority of their questions and difficulties can be resolved in the group, or by consultation with another group. I do answer for the class those few questions that none of the groups have been able to handle.

Talk to your friends at other institutions about the course. Despite the lack of public discussion and major reforms, there are a variety of innovative ideas being tried by instructors all over the map. In two conversations at the math meetings, for example, I learned of a "lab manual" and a plastic device for visualizing reflections in the plane that I will be able to adapt to my own course.

Don't let your students (or yourself) get bored. In my experience this is quite likely to happen in the early part of the course, which deals with integer and whole number arithmetic, topics that form the bulk of the mathematics in the primary grades. Even students with weak mathematical backgrounds (often the majority of the class) know the *mechanics* of addition, subtraction, multiplication, and division. Having mastered rote procedures for calculating, they pay little attention to *why* the various algorithms work or how they will explain them to children.

I have found that presenting this material in base five, without telling the students that it *is* base five keeps me awake and jars the students out of complacency. The key is to present the material, not as a "translation exercise" from base ten, but as a "new" arithmetic in which you count 1, 2, 3, 4, 10 (fen), 11 (fenone), 12 (fentwo), . . . , 21 (twofen-one), . . . , 100 (one fundred), etc. I tell them to forget all the math they learned in the past—that we are going to recreate a child's mathematical experiences in the primary grades. This new "fen arithmetic" is just similar enough to the "old" one that students aren't overly anxious and just different enough that *they can't rely on what they learned by rote in the past*; they now must think about what's going on. For full details, including student reactions, see [2].

Undoubtedly, others who have taught the course can contribute other or better suggestions—and I certainly hope they will.

REFERENCES

1. H. Clemens, The Park City Institute: A Mathematician's Apology, *Notices of the AMS* 39 (March 1992).
2. T. W. Hungerford, An Experiment in Teaching Prospective Elementary School Teachers, *UME TRENDS* 4 (March 1992).

3. J. R. C. Leitzel, Preparing Elementary School Mathematics Teachers, *UME TRENDS* 3 (December 1991).
4. Mathematical Association of America, *A Call for Change: Recommendations for the Mathematical Preparation of Teachers*, MAA, Washington, D.C., 1991.
5. National Council of Teachers of Mathematics, *Curriculum and Evaluation Standards for School Mathematics*, NCTM, Reston, VA, 1989.

*Department of Mathematics
Cleveland State University
Cleveland, OH 44115*

LETTER TO THE EDITOR

The theorem of Abel, given an elegant new proof by Grosz and Taiani in the June-July issue of v. 100 (1993) (575–576), could also be proved by the Lagrange Interpolation formula applied to $P(r)$ at the points $\{r_1, \dots, r_n\}$, by equating the coefficient of $r^{(n-1)}$ on both sides. The Lagrange Interpolation formula, in turn, relies on the fact that a polynomial of degree n is uniquely determined by its values at $n + 1$ distinct points, that is the “standard” application of Vandermonde’s determinant mentioned at the first paragraph.

Doron Zeilberger
Department of Mathematics
Temple University
Philadelphia, PA 19122

Galois Theory for Beginners

John Stillwell

Galois theory is rightly regarded as the peak of undergraduate algebra, and the modern algebra syllabus is designed to lead to its summit, usually taken to be the unsolvability of the general quintic equation. I fully agree with this goal, but I would like to point out that most of the equipment supplied—in particular normal extensions, irreducible polynomials, splitting fields and a lot of group theory—is unnecessary. The biggest encumbrance is the so-called “fundamental theorem of Galois theory.” This theorem, interesting though it is, has little to do with polynomial equations. It relates the subfield structure of a normal extension to the subgroup structure of its group, and can be proved without use of polynomials (see, e.g., the appendix to Tignol [6]). Conversely, one can prove the unsolvability of polynomial equations without knowing about normality of field extensions or the Galois correspondence between subfields and subgroups.

The aim of this paper is to prove the unsolvability by radicals of the quintic (in fact of the general n th degree equation for $n \geq 5$) using just the fundamentals of groups, rings and fields from a standard first course in algebra. The main fact it will be necessary to know is that if ϕ is a homomorphism of group G onto group G' then $G' \cong G/\ker \phi$, and conversely, if $G/H \cong G'$ then H is the kernel of a homomorphism of G onto G' . The concept of Galois group, which guides the whole proof, will be defined when it comes up. With this background, a proof of unsolvability by radicals can be constructed from just three basic ideas, which will be explained more fully below:

1. Fields containing n indeterminates can be “symmetrized”.
2. The Galois group of a radical extension is solvable.
3. The symmetric group S_n is not solvable.

When one considers the number of mathematicians who have worked on Galois theory, it is not possible to believe this proof is really new. In fact, all proofs seem to contain steps similar to the three just listed. Nevertheless, most of the standard approach had to be stripped away before the present proof became visible. I read the books of Edwards [2], Tignol [6], Artin [1], Kaplansky [3], MacLane and Birkhoff [5] and Lang [4], taught a course in Galois theory, and then discarded 90% of what I had learned.

I wish to thank my students, particularly Mark Kisin, for helpful suggestions and discussions which led to the writing of this paper. I am also grateful to the referee for several improvements.

THE GENERAL EQUATION OF DEGREE n . The goal of classical algebra was to express the roots of the general n th degree equation

$$(*) \quad x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = 0$$

in terms of the coefficients a_0, \dots, a_{n-1} , using a finite number of operations $+, -, \times, \div$ and radicals $\sqrt{}, \sqrt[3]{}, \dots$. For example, the roots x_1, x_2 of the general quadratic equation

$$x^2 + a_1x + a_0 = 0$$

are expressed by the formula

$$x_1, x_2 = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_0}}{2}.$$

Formulas for the roots of general cubic and quartic equations are also known, using cube roots as well as square roots. We say that these equations are *solvable by radicals*.

The set of elements obtainable from a_0, \dots, a_{n-1} by $+, -, \times, \div$ is the *field* $\mathbb{Q}(a_0, \dots, a_{n-1})$. If we denote the roots of (*) by x_1, \dots, x_n , so that

$$(x - x_1) \cdots (x - x_n) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$$

then a_0, \dots, a_{n-1} are polynomial functions of x_1, \dots, x_n called the *elementary symmetric functions*:

$$a_0 = (-1)^n x_1 x_2 \cdots x_n, \dots, a_{n-1} = -(x_1 + x_2 + \cdots + x_n).$$

The goal of solution by radicals is then to *extend* $\mathbb{Q}(a_0, \dots, a_{n-1})$ by *adjoining radicals* until a field containing the roots x_1, \dots, x_n is obtained. For example, the roots x_1, x_2 of the quadratic equation lie in the extension of $\mathbb{Q}(a_0, a_1) = \mathbb{Q}(x_1 x_2, x_1 + x_2)$ by the radical

$$\sqrt{a_1^2 - 4a_0} = \sqrt{(x_1 + x_2)^2 - 4x_1 x_2} = \sqrt{(x_1 - x_2)^2} = \pm(x_1 - x_2).$$

In this case we get $\mathbb{Q}(x_1, x_2)$ itself as the radical extension $\mathbb{Q}(a_0, a_1, \sqrt{a_1^2 - 4a_0})$, though in other cases a radical extension of $\mathbb{Q}(a_0, \dots, a_{n-1})$ containing x_1, \dots, x_n is larger than $\mathbb{Q}(x_1, \dots, x_n)$. In particular, the solution of the cubic equation gives a radical extension of $\mathbb{Q}(a_0, a_1, a_2)$ which includes imaginary cube roots of unity as well as x_1, x_2, x_3 .

In general, adjoining an element α to a field F means forming the closure of $F \cup \{\alpha\}$ under $+, -, \times, \div$ (by a non-zero element), i.e., taking the intersection of all fields containing $F \cup \{\alpha\}$. The adjunction is called *radical* if some positive integer power α^m of α equals an element $f \in F$, in which case α may be represented by the radical expression $\sqrt[m]{f}$. The result $F(\alpha_1)(\alpha_2) \dots (\alpha_k)$ of successive adjunctions is denoted by $F(\alpha_1, \dots, \alpha_k)$ and if each adjunction is radical we say $F(\alpha_1, \dots, \alpha_k)$ is a *radical extension* of F .

It is clear from these definitions that a radical extension E of $\mathbb{Q}(a_0, \dots, a_{n-1})$ containing x_1, \dots, x_n is also a radical extension of $\mathbb{Q}(x_1, \dots, x_n)$, since $a_0, \dots, a_{n-1} \in \mathbb{Q}(x_1, \dots, x_n)$. Thus we also have to study radical extensions of $\mathbb{Q}(x_1, \dots, x_n)$. The most important property of $\mathbb{Q}(x_1, \dots, x_n)$ is that it is *symmetric* with respect to x_1, \dots, x_n , in the sense that any permutation σ of x_1, \dots, x_n extends to a bijection σ of $\mathbb{Q}(x_1, \dots, x_n)$ defined by

$$\sigma f(x_1, \dots, x_n) = f(\sigma x_1, \dots, \sigma x_n)$$

for each rational function f of x_1, \dots, x_n . Moreover, this bijection σ obviously

satisfies

$$\sigma(f + g) = \sigma f + \sigma g,$$

$$\sigma(fg) = \sigma f \cdot \sigma g,$$

and hence is an *automorphism* of $\mathbb{Q}(x_1, \dots, x_n)$.

A radical extension E of $\mathbb{Q}(x_1, \dots, x_n)$ is not necessarily symmetric in this sense. For example, $\mathbb{Q}(x_1, \dots, x_n, \sqrt{x_1})$ contains a square root of x_1 , but not of x_2 , hence there is no automorphism exchanging x_1 and x_2 . However, we can restore symmetry by adjoining $\sqrt{x_2}, \dots, \sqrt{x_n}$ as well. The obvious generalization of this idea gives a way to “symmetrize” any radical extension E of $\mathbb{Q}(x_1, \dots, x_n)$:

Theorem 1. *For each radical extension E of $\mathbb{Q}(x_1, \dots, x_n)$ there is a radical extension $\bar{E} \supseteq E$ with automorphisms σ extending all permutations of x_1, \dots, x_n .*

Proof: For each adjoined element, represented by radical expression $e(x_1, \dots, x_n)$, and each permutation σ of x_1, \dots, x_n , adjoin the element $e(\sigma x_1, \dots, \sigma x_n)$. Since there are only finitely many permutations σ , the resulting field $\bar{E} \supseteq E$ is also a radical extension of $\mathbb{Q}(x_1, \dots, x_n)$.

This gives a bijection (also called σ) of \bar{E} sending each $f(x_1, \dots, x_n) \in \bar{E}$ (a rational function of x_1, \dots, x_n and the adjoined radicals) to $f(\sigma x_1, \dots, \sigma x_n)$, and this bijection is obviously an automorphism of \bar{E} , extending the permutation σ . ■

The reason for wanting an automorphism σ extending each permutation of x_1, \dots, x_n is that a_0, \dots, a_{n-1} are fixed by such permutations, and hence so is every element of the field $\mathbb{Q}(a_0, \dots, a_{n-1})$. If $E \supseteq F$ are any fields, the automorphisms σ of E fixing all elements of F form what is called the *Galois group of E over F* , $\text{Gal}(E/F)$. This concept alerts us to the following corollary of Theorem 1:

Corollary. *If E is a radical extension of $\mathbb{Q}(a_0, \dots, a_{n-1})$ containing x_1, \dots, x_n then there is a further radical extension $\bar{E} \supseteq E$ such that $\text{Gal}(\bar{E}/\mathbb{Q}(a_0, \dots, a_{n-1}))$ includes automorphisms σ extending all permutations of x_1, \dots, x_n .*

Proof: This is immediate from Theorem 1 and the fact that a radical extension of $\mathbb{Q}(a_0, \dots, a_{n-1})$ containing x_1, \dots, x_n is also a radical extension of $\mathbb{Q}(x_1, \dots, x_n)$. ■

THE STRUCTURE OF RADICAL EXTENSIONS. So far we know that a solution by radicals of the general n th degree equation (*) entails a radical extension of $\mathbb{Q}(a_0, \dots, a_{n-1})$ containing x_1, \dots, x_n , and hence a radical extension \bar{E} with the symmetry described in the corollary above. This opens a route to prove *non-existence* of such a solution by learning enough about $\text{Gal}(\bar{E}/\mathbb{Q}(a_0, \dots, a_{n-1}))$ to show that such symmetry is lacking, at least for $n \geq 5$. In the present section we shall show that the Galois group $\text{Gal}(F(\alpha_1, \dots, \alpha_k)/F)$ of any radical extension has a special structure, called *solvability*, inherited from the structure of $F(\alpha_1, \dots, \alpha_k)$. Then in the next section we shall show that this structure is indeed incompatible with the symmetry described in the corollary. To simplify the derivation of this structure, we shall show that certain assumptions about the adjunction of radicals α_i can be made without loss of generality.

First, we can assume that each radical α_i adjoined is a p th root for some *prime* p . E.g., instead of adjoining $\sqrt[6]{\alpha}$ we can adjoin first $\sqrt{\alpha} = \beta$, then $\sqrt[3]{\beta}$. Second, if α_i

is a p th root we can assume that $F(\alpha_1, \dots, \alpha_i)$ contains no p th roots of unity not in $F(\alpha_1, \dots, \alpha_{i-1})$ unless α_i itself is a p th root of unity. If this is not the case initially we simply adjoin a p th root of unity $\zeta \neq 1$ to $F(\alpha_1, \dots, \alpha_{i-1})$ before adjoining α_i (in which case $F(\alpha_1, \dots, \alpha_{i-1}, \zeta)$ contains all the p th roots of unity: $1, \zeta, \zeta^2, \dots, \zeta^{p-1}$). With both these modifications the final field $F(\alpha_1, \dots, \alpha_k)$ is the same, and it remains the same if the newly adjoined roots ζ are included in the list $\alpha_1, \dots, \alpha_k$. Hence we have:

Any radical extension $F(\alpha_1, \dots, \alpha_k)$ is the union of an ascending tower of fields

$$F = F_0 \subseteq F_1 \subseteq \dots \subseteq F_k = F(\alpha_1, \dots, \alpha_k)$$

where each $F_i = F_{i-1}(\alpha_i)$, α_i is the p_i -th root of an element in F_{i-1} , p_i is prime, and F_i contains no p_i -th roots of unity not in F_{i-1} unless α_i is itself a p_i -th root of unity.

Corresponding to this tower of fields we have a descending tower of groups

$$\text{Gal}(F_k/F_0) = G_0 \supseteq G_1 \supseteq \dots \supseteq G_k = \text{Gal}(F_k/F_k) = \{1\}$$

where $G_i = \text{Gal}(F_k/F_i) = \text{Gal}(F_k/F_{i-1}(\alpha_i))$ and 1 denotes the identity automorphism. The containments are immediate from the definition of $\text{Gal}(E/B)$, for any fields $E \supseteq B$, as the group of automorphisms of E fixing each element of B . As B increases to E , $\text{Gal}(E/B)$ must decrease to $\{1\}$. The important point is that the step from G_{i-1} to its subgroup G_i , reflecting the adjunction of the p_i -th root α_i to F , is “small” enough to be describable in group-theoretic terms: G_i is a normal subgroup of G_{i-1} , and G_{i-1}/G_i is abelian, as we shall now show.

To simplify notation further we set

$$E = F_k, B = F_{i-1}, \alpha = \alpha_i, p = p_i,$$

so the theorem we want is:

Theorem 2. *If $E \supseteq B(\alpha) \supseteq B$ are fields with $\alpha^p \in B$ for some prime p , and if $B(\alpha)$ contains no p th roots of unity not in B unless α itself is a p th root of unity, then $\text{Gal}(E/B(\alpha))$ is a normal subgroup of $\text{Gal}(E/B)$ and $\text{Gal}(E/B)/\text{Gal}(E/B(\alpha))$ is abelian.*

Proof: By the homomorphism theorem for groups, it suffices to find a homomorphism of $\text{Gal}(E/B)$, with kernel $\text{Gal}(E/B(\alpha))$, into an abelian group (i.e., onto a subgroup of an abelian group, which of course is also abelian). The obvious map with kernel $\text{Gal}(E/B(\alpha))$ is *restriction to $B(\alpha)$* , $\sigma|_{B(\alpha)}$, since by definition

$$\sigma \in \text{Gal}(E/B(\alpha)) \Leftrightarrow \sigma|_{B(\alpha)} \text{ is the identity map.}$$

The homomorphism property,

$$\sigma'\sigma|_{B(\alpha)} = \sigma'|_{B(\alpha)}\sigma|_{B(\alpha)} \quad \text{for all } \sigma', \sigma \in \text{Gal}(E/B),$$

is automatic provided $\sigma|_{B(\alpha)}(b) \in B(\alpha)$ for each $b \in B(\alpha)$, i.e. provided $B(\alpha)$ is closed under each $\sigma \in \text{Gal}(E/B)$.

Since σ fixes B , $\sigma|_{B(\alpha)}$ is completely determined by the value $\sigma(\alpha)$. If α is a p th root of unity ζ then

$$(\sigma(\alpha))^p = \sigma(\alpha^p) = \sigma(\zeta^p) = \sigma(1) = 1,$$

hence $\sigma(\alpha) = \zeta^i = \alpha^i \in B(\alpha)$, since each p th root of unity is some ζ^i . If α is not a

root of unity then

$$(\sigma(\alpha))^p = \sigma(\alpha^p) = \alpha^p \quad \text{since } \alpha^p \in B,$$

hence $\sigma(\alpha) = \zeta^j \alpha$ for some p th root of unity ζ , and $\zeta \in B$ by hypothesis, so again $\sigma(\alpha) \in B(\alpha)$. Thus $B(\alpha)$ is closed as required.

This also implies that $|_{B(\alpha)}$ maps $\text{Gal}(E/B)$ into $\text{Gal}(B(\alpha)/B)$, so it now remains to check that $\text{Gal}(B(\alpha)/B)$ is abelian. If α is a root of unity then, as we have just seen, each $\sigma|_{B(\alpha)} \in \text{Gal}(B(\alpha)/B)$ is of the form σ_i , where $\sigma_i(\alpha) = \alpha^i$, hence

$$\sigma_i \sigma_j(\alpha) = \sigma_i(\alpha^j) = \alpha^{ij} = \sigma_j \sigma_i(\alpha).$$

Likewise, if α is not a root of unity then each $\sigma|_{B(\alpha)} \in \text{Gal}(B(\alpha)/B)$ is of the form σ_i where $\sigma_i(\alpha) = \zeta^i \alpha$, hence

$$\sigma_i \sigma_j(\alpha) = \sigma_i(\zeta^j \alpha) = \zeta^{i+j} \alpha = \sigma_j \sigma_i(\alpha)$$

since $\zeta \in B$ and therefore ζ is fixed. Hence in either case $\text{Gal}(B(\alpha)/B)$ is abelian. ■

The property of $\text{Gal}(F(\alpha_1, \dots, \alpha_k)/F)$ implied by this theorem, that it has subgroups $\text{Gal}(F(\alpha_1, \dots, \alpha_k)/F) = G_0 \supseteq G_1 \supseteq \dots \supseteq G_k = \{1\}$ with each G_i normal in G_{i-1} and G_{i-1}/G_i abelian, is called *solubility* of $\text{Gal}(F(\alpha_1, \dots, \alpha_k)/F)$.

NON-EXISTENCE OF SOLUTIONS BY RADICALS WHEN $n \geq 5$. As we have said, this amounts to proving that a radical extension of $\mathbb{Q}(a_0, \dots, a_{n-1})$ does not contain x_1, \dots, x_n or, equivalently, $\mathbb{Q}(x_1, \dots, x_n)$. We have now reduced this problem to proving that the symmetry of the hypothetical extension \bar{E} containing x_1, \dots, x_n , given by the corollary to Theorem 1, is incompatible with the solubility of $\text{Gal}(\bar{E}/\mathbb{Q}(a_0, \dots, a_{n-1}))$, given by Theorem 2. Our proof looks only at the effect of the hypothetical automorphisms of \bar{E} on x_1, \dots, x_n , and hence it is really about the *symmetric group* S_n of all permutations of x_1, \dots, x_n . In fact, we are adapting a standard proof that S_n is not a solvable group, given by Milgram in his appendix to Artin [1].

Theorem 3. *A radical extension of $\mathbb{Q}(a_0, \dots, a_{n-1})$ does not contain $\mathbb{Q}(x_1, \dots, x_n)$ when $n \geq 5$.*

Proof: Suppose on the contrary that E is a radical extension of $\mathbb{Q}(a_0, \dots, a_{n-1})$ which contains $\mathbb{Q}(x_1, \dots, x_n)$. Then E is also a radical extension of $\mathbb{Q}(x_1, \dots, x_n)$ and by the corollary to Theorem 1 there is a radical extension $\bar{E} \supseteq E$ such that $G_0 = \text{Gal}(\bar{E}/\mathbb{Q}(a_0, \dots, a_{n-1}))$ includes automorphisms σ extending all permutations of x_1, \dots, x_n .

By Theorem 2, G_0 has a decomposition

$$G_0 \supseteq G_1 \supseteq \dots \supseteq G_k = \{1\}$$

where each G_{i+1} is a normal subgroup of G_i and G_{i-1}/G_i is abelian. We now show that this is contrary to the existence of the automorphisms σ .

Since G_{i-1}/G_i is abelian, G_i is the kernel of a homomorphism of G_{i-1} onto an abelian group, and therefore

$$\sigma, \tau \in G_{i-1} \Rightarrow \sigma^{-1} \tau^{-1} \sigma \tau \in G_i.$$

We use this fact to prove by induction on i that, if $n \geq 5$, each G_i contains automorphisms σ extending all 3-cycles (x_a, x_b, x_c) . This is true for G_0 by

hypothesis, and when $n \geq 5$ the property persists from G_{i-1} to G_i because

$$(x_a, x_b, x_c) = (x_d, x_a, x_c)^{-1}(x_c, x_e, x_b)^{-1}(x_d, x_a, x_c)(x_c, x_e, x_b)$$

where a, b, c, d, e are distinct. Thus if there are at least five indeterminates x_j , there are σ in each G_i which extend arbitrary 3-cycles (x_a, x_b, x_c) , and this means in particular that $G_k \neq \{1\}$. This contradiction shows that $\mathbb{Q}(x_1, \dots, x_n)$ is not contained in any radical extension of $\mathbb{Q}(a_0, \dots, a_{n-1})$ when $n \geq 5$. ■

REFERENCES

1. E. Artin, *Galois Theory*, Notre Dame, 1965.
2. H. M. Edwards, *Galois Theory*, Springer-Verlag, New York, 1984.
3. I. Kaplansky, *Fields and Rings*, University of Chicago Press, 1969.
4. S. Lang, *Undergraduate Algebra*, Springer-Verlag, New York, 1987.
5. S. MacLane & G. Birkhoff, *Algebra*, 2nd ed, Collier Macmillan, New York, 1979.
6. J.-P. Tignol, *Galois' Theory of Algebraic Equations*, Longman, New York, 1988.

Department of Mathematics
 Monash University
 Clayton 3168
 Australia

PICTURE PUZZLE
(from the collection of Paul Halmos)



This famous topologist was usually
 considered more scary than scared.
 (see page 86.)

A Family of Invertible Matrices

10214 [1992, 362]. *Proposed by Stephen Penrice, Emory University, Atlanta, GA.*

For all integers $n > 1$, let $f(n)$ denote the largest real number such that, for any set of non-negative real numbers satisfying $a_1 + \cdots + a_n < f(n)$, the n by n matrix with a_1, \dots, a_n along the main diagonal and -1 in all other positions is invertible. Show that $f(n)$ is well-defined, and obtain an explicit formula for it.

Solution by Dirk P. Laurie, Potchefstroom University for Christian Higher Education, Vanderbijlpark, South Africa. The matrix concerned can be written as $\mathbf{M} = \mathbf{D} - \mathbf{e}\mathbf{e}^T$, where $\mathbf{D} = \text{diag}(1 + a_i)$ and \mathbf{e} is a column of ones. By The Sherman-Morrison formula (see Alston S. Householder, *The Theory of Matrices in Numerical Analysis*, Dover, 1964, sect. 5.1),

$$\mathbf{M}^{-1} = \mathbf{D}^{-1}(1 - \mathbf{e}^T \mathbf{D}^{-1} \mathbf{e})^{-1} \mathbf{D}^{-1} \mathbf{e} \mathbf{e}^T \mathbf{D}^{-1};$$

hence \mathbf{M} is invertible whenever $\mathbf{e}^T \mathbf{D}^{-1} \mathbf{e} \neq 1$. Using the fact that the arithmetic mean of positive numbers is not less than their harmonic mean, we obtain

$$\begin{aligned} \mathbf{e}^T \mathbf{D}^{-1} \mathbf{e} &= \sum_{i=1}^n \frac{1}{1 + a_i} \\ &\geq n \left(\sum_{i=1}^n \frac{1 + a_i}{n} \right)^{-1} \\ &= \frac{n^2}{n + f(n)} \\ &> 1 \text{ if } f(n) < n^2 - n. \end{aligned}$$

The function $f(n) = n^2 - n$ is best possible, since if all $a_i = n - 1$, then \mathbf{M} is not invertible because $\mathbf{M}\mathbf{e} = \mathbf{0}$.

Editorial comment. Many solvers noted that the same result holds if the condition $a_i \geq 0$ is weakened to $a_i \geq -1$.

Solved also by J. C. Binz (Switzerland), B. W. Brock, F. Brulois and T. Shore and D. B. Tyler, D. Callan, R. J. Chapman (U.K.), M. Dindos (Slovakia), Z. Franco, K. S. Kedlaya (student), N. Komanda, O. Krafft (Germany), O. P. Lossers (The Netherlands), M. Mócsy (Hungary), A. Nijenhuis, A. Pedersen (Denmark), E. Schmeichel, R. Stong, M. Tsatsomeris (Canada), and the proposer.

Collaborating editors: David F. Appleyard, Paul T. Bateman, Duane M. Broline, Barry W. Brunson, Frank S. Cater, Gulbank D. Chakerian, Underwood Dudley, Gerald A. Edgar, Michael A. Filaseta, Ira M. Gessel, Richard A. Gibbs, Jerrold R. Griggs, Douglas A. Hensley, John R. Isbell, Mourad E. H. Ismail, Murray Klamkin, Daniel J. Kleitman, Frederick W. Luttman, Frank B. Miles, Richard Pfiefer, Stephen L. Portnoy, J. O. Shallit, John Henry Steelman, Kenneth B. Stolarsky, David E. Tepper, Douglas B. Tyler, Daniel Ullman, and William E. Watkins.

Answer to Picture Puzzle

(p. 27)

Solomon Lefschetz.

Into The Hourglass: Reflections on the Forces Acting on a Granular Material

E. Bruce Pitman

The storage and flow of granular materials is common to many industries. For example, coal is stored in large bunkers at electric generating plants, to be used later in fueling the plant furnaces. A very different kind of material storage occurs at cereal manufacturing facilities; here the corn flakes are baked, stored in large bins temporarily, and then packaged. A natural question to ask is How strong do the storage bins have to be? Engineers would like to be sure that the coal bunkers do not collapse under the weight of the coal (as sometimes happens, often with loss of life). At the same time, the cereal manufacturer wants to be sure that he doesn't crush the corn flakes into tiny bits before packaging. How does one model the forces acting in a granular medium? What kind of pressures build up inside a vessel storing granular materials? These are the questions we address in this article.

It would be natural to examine each particle individually, resolve all the forces acting on each particle, and then solve Newton's law $\text{Force} = \text{Mass} \cdot \text{Acceleration}$ for each particle. Needless to say, with perhaps 10^8 or more particles in a bin, such a procedure would be prohibitively difficult to carry out. We make a simplification here. In trying to determine forces inside the bin, we forget the particle nature of the material and treat it as a continuum. Once we make this simplification, we are faced with the issue of describing the forces in a continuum, in particular of describing the constitutive behavior of the material [that is, explaining the stress-strain relationship which models the continuum]. We examine a simple constitutive model later on.

Before looking at the pressures inside a bin of granular material, it would be useful to recall the analogous question for fluids. How much pressure is there at the bottom of a tall water tower? A look at your basic physics textbook will tell you that the pressure at a point at the bottom of a water tower equals the weight of all the water above that point. The density of water is 1 gram/cm^3 ; you should calculate the pressure at the base of a water tower comprised of a cylindrical tower, perhaps 20 meters in diameter and 30 meters high, with a spherical top of perhaps 20 meters radius.

In order to estimate the pressure inside a bin of granular material, such as illustrated in FIGURE 2, we need to introduce the concept of stress; this is the subject of the next section. We then introduce a generalization of sliding friction laws as a constitutive model for granular material. Finally we look at the forces inside a cylindrical bin and a converging hopper.

In the sections to follow, we restrict attention to two dimensions; an analysis can be carried out in three dimensions, but the 2-D case is a little easier to work out.

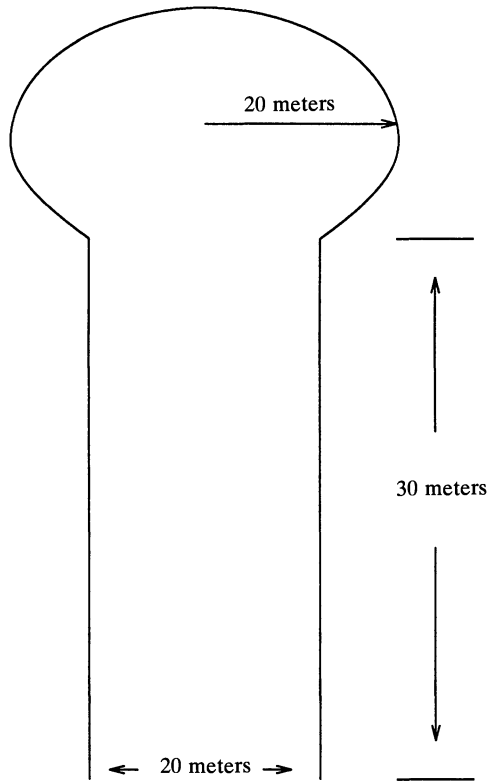


Figure 1. A schematic diagram of a watertower.

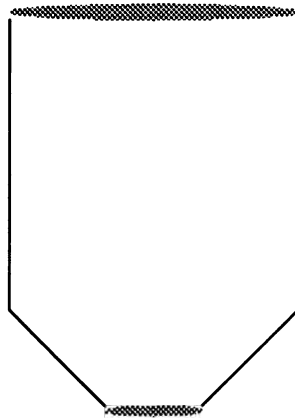


Figure 2. A typical bin, with a cylindrical part and a converging hopper.

WHAT IS STRESS? In this section we introduce the concept of stress, and generalize the idea of sliding friction to model the deformation of granular materials. Let us begin with stress. We may take as a working definition of stress the following:

Definition. Stress is the internal force in a continuous medium acting across a boundary.

To explain this definition a bit, imagine a region $\Omega \subset \mathbb{R}^2$. If F^i is the sum of all external forces in the i th direction acting on Ω , per unit volume, then we write

$$\int_{\Omega} F^i dV = \int_{\partial\Omega} \sigma^i(\mathbf{x}, \nu) ds. \quad (1)$$

Here dV is the volume element for Ω and ds is the area element on $\partial\Omega$, the boundary of Ω . The integrand $\sigma^i(\mathbf{x}, \nu)$ is called the surface traction in the i th direction; it depends on position \mathbf{x} and also upon the outward normal ν to the boundary.

Cauchy showed that the dependence of the traction on ν is linear; that is,

$$\sigma^i(\mathbf{x}, \nu) = \sigma^{ij}(\mathbf{x}) \nu^j(\mathbf{x}).$$

The stress σ^{ij} gives the force per unit area acting in the i th direction across a surface whose normal is in the j th direction. Using this relation, we substitute into Eq. (1) to get

$$\int_{\Omega} F^i dV = \int_{\partial\Omega} \sigma^{ij}(\mathbf{x}) \nu^j(\mathbf{x}) ds.$$

Then applying the Gauss Theorem, we have

$$\int_{\Omega} F^i dV = \int_{\Omega} \frac{\partial}{\partial x_j} \sigma^{ij}(\mathbf{x}) dV.$$

Since the region Ω is arbitrary, the integrands must be equal:

$$F^i = \frac{\partial}{\partial x_j} \sigma^{ij}. \quad (2)$$

Equation 2 expresses the balance of momentum, or force equilibrium; it says that the forces acting in the i th direction are balanced by the gradient of the stress.

The stresses σ^{ij} are elements of a matrix, Σ . If we call the 1-direction x and the 2-direction y , then

$$\Sigma = \begin{pmatrix} \sigma^{xx} & \sigma^{xy} \\ \sigma^{xy} & \sigma^{yy} \end{pmatrix}.$$

In writing this matrix, we have used the fact that $\sigma^{xy} = \sigma^{yx}$, which arises from conservation of angular momentum.

You are already familiar with certain kinds of stress. For example air pressure is a stress, albeit a stress of a simple form. Air pressure acts with equal magnitude in all directions simultaneously; as a matrix, it looks like a multiple of the identity matrix. In our examination of granular materials, the off-diagonal entries will be important.

It is common in granular mechanics to consider compressive stresses as positive. This is natural for granular materials, since any tensile stress would pull the material apart causing it to disintegrate. This convention will introduce an extra minus sign later on that we must be aware of. In other branches of mechanics, the convention is different.

We now need to model the deformation of a granular material. If the deformation process is slow, then as the particles within the medium move, they slide over each other, one particle remaining in contact with another for an extended period of time. A simple model of such behavior is a generalization of sliding friction. Recall that the normal and tangential forces, N and T , acting on a block which sits on an inclined plane are related $T \leq \mu^b N$. The block slides (i.e. deformation

occurs) only when equality holds. The friction coefficient μ^b is often related to the angle ψ which the plane makes with the horizontal— $\mu^b = \tan(\psi)$. How can we generate a model for deformation of a granular material based on similar ideas? We must generalize the notions of tangential and normal force.

Since Σ is a 2×2 matrix, it has two eigenvalues. Let call these σ^1 and σ^2 , with $\sigma^1 \geq \sigma^2$. Recall from Linear Algebra that the determinant and the trace of two similar matrices are the same. [The trace of a matrix is the sum of its diagonal elements.] Using this fact, let us decompose Σ as follows:

$$\begin{pmatrix} \sigma^{xx} & \sigma^{xy} \\ \sigma^{xy} & \sigma^{yy} \end{pmatrix} = \frac{\sigma^{xx} + \sigma^{yy}}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \frac{\sigma^{xx} - \sigma^{yy}}{2} & \sigma^{xy} \\ \sigma^{xy} & -\frac{\sigma^{xx} - \sigma^{yy}}{2} \end{pmatrix}. \quad (3)$$

Now make a similarity transformation, conjugating Σ by a change of basis matrix, call it P , whose columns are the right eigenvectors associated to the eigenvalues σ^1 and σ^2 . We can again decompose this new matrix into the form

$$P^{-1}\Sigma P = \frac{\sigma^1 + \sigma^2}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \frac{\sigma^1 - \sigma^2}{2} & 0 \\ 0 & -\frac{\sigma^1 - \sigma^2}{2} \end{pmatrix}. \quad (4)$$

The pressure-like term $(\sigma^1 + \sigma^2)/2$ plays the role of the normal force. The tangential force is given by the term $(\sigma^1 - \sigma^2)/2$. Sliding friction can then be generalized to read

$$\frac{\sigma^1 - \sigma^2}{2} = \mu \frac{\sigma^1 + \sigma^2}{2}. \quad (5)$$

Equation (5) is written in terms of the eigenvalues σ^1 and σ^2 ; as an exercise, express this equation in terms of σ^{xx} , σ^{yy} , and σ^{xy} .

What should we use as the friction coefficient μ ? A simple solution suggests itself if we think about everyone's favorite granular material, sand. When you were younger, perhaps you had a sandbox in your backyard. You might have tried to build castles and forts in the sand. If so, you undoubtedly noticed that you can make a heap of sand, the sides of which make a certain angle with the horizontal. If you tried to make the heap steeper, you couldn't. A sandslide ensued and the heap retained its characteristic angle. That angle is called the angle of internal friction, often denoted by ϕ , and plays a role analogous to the inclined plane angle ψ . FIGURE 3 illustrates this relation.

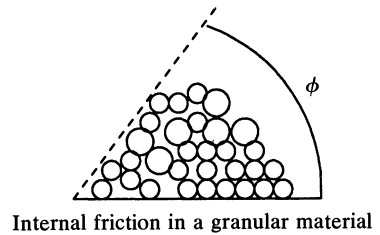
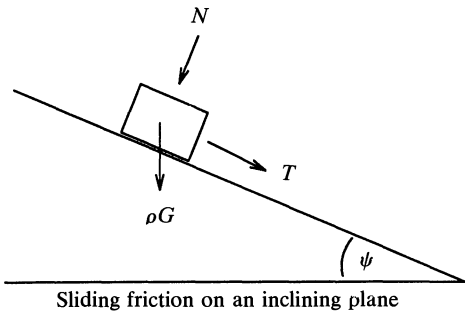


Figure 3.

When a granular material deforms, the particles can slide along any plane they wish and are not restricted to slide along a particular plane. Thus sliding will occur along whichever plane first mobilizes sufficient forces to be in the ratio given by Equation (5). In FIGURE 4, we have drawn a diagram showing Normal and Tangential forces, and a circle whose center is at $(\sigma^1 + \sigma^2)/2$ and whose radius is $(\sigma^1 - \sigma^2)/2$. This so-called Mohr Circle defines the directions in which Equation 5 first allows frictional sliding. Drawing the angle ϕ , the usual constraint of sliding friction imposes the condition $T = \mu N$. Simple trigonometry then gives a relation between the eigenvalues of Σ

$$\sin(\phi) = \frac{\sigma^1 - \sigma^2}{\sigma^1 + \sigma^2}. \quad (6)$$

Thus the friction coefficient μ in (5) is equal to $\sin(\phi)$.

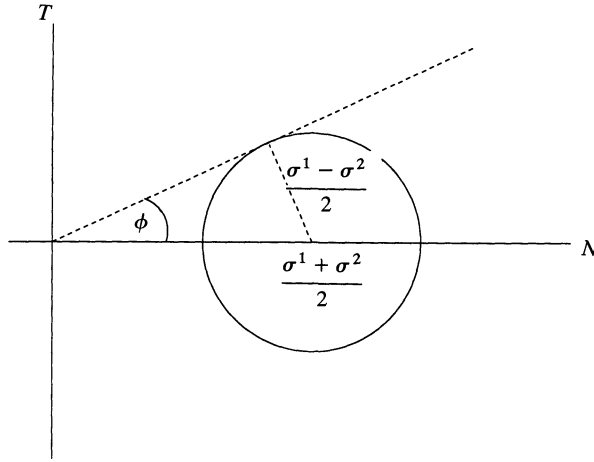


Figure 4. Resolution of Normal and Tangential forces acting in a granular material.

Before leaving this section, we remark that one critical difference between granular materials and fluids is the internal friction angle. For a material like sand, $\phi \approx 35^\circ$. However, you cannot make a heap of fluid like you can with sand; fluids have an angle of internal friction of 0° . This difference dramatically alters the stress distribution set up within a sample of fluid or granular material.

FORCES IN A BIN. We are now in a position to compute the forces that act on a bin holding granular material. We will begin by considering the simpler case of forces inside a cylindrical vessel, and then compute the forces in a converging hopper. In the following, we consider the static forces built up when a granular material rests inside a bin; inclusion of dynamic effects adds significant complexity to the problem.

Cylindrical bin. Consider the forces shown in FIGURE 5, where ρ is the density of the material and G is the acceleration of gravity. In the diagram, we have not explicitly shown the internal stresses. Assume the origin of a Cartesian coordinate system lies along the centerline of the bin along the line marked $y = 0$. Instead of a complete analysis, we content ourselves with an approximate analysis, examining

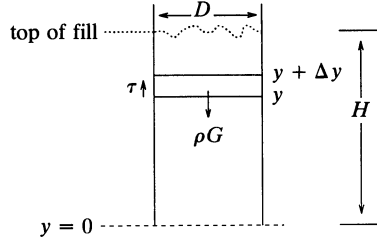


Figure 5. Force balance for a cylindrical bin.

the average stress

$$\bar{\sigma} = \int_{-D/2}^{D/2} \sigma^{yy}(x, y) dx.$$

Let τ be the frictional force mobilized along the bin wall. Then force balance requires

$$-D[\bar{\sigma}(y + \Delta y) - \bar{\sigma}(y)] + 2\tau\Delta y = \rho G\Delta y D.$$

The extra minus sign on the first term comes from the convention of treating compressive stresses as positive. If we divide through by $D\Delta y$ and let $\Delta y \rightarrow 0$ we have the equilibrium equation

$$-\frac{\partial}{\partial y}\bar{\sigma} + \frac{2\tau}{D} = \rho G. \quad (7)$$

We now must model the wall force τ . Again our fundamental idea comes from simple sliding friction. The stress τ is a shearing force, akin to the tangential force in the inclined plane figure; in our stress notation, τ is the σ^{xy} stress. We assume this stress is equal to a friction coefficient times the “normal” stress σ^{xx} :

$$\sigma^{xy}(D/2, y) = \eta\sigma^{xx}(D/2, y).$$

From the sliding friction model, we assume the wall friction coefficient $\eta = \tan(\phi_w)$, where ϕ_w is the wall-material friction angle. Also, we assume that the eigendirection corresponding to σ^1 is parallel to the x -axis, and the eigendirection corresponding to σ^2 is parallel to the y -axis. Then σ^{xx} and σ^{yy} are the eigenvalues, and

$$\sigma^{xx}(x, y) = K\sigma^{yy}(x, y)$$

where $K = (1 + \sin(\phi))/(1 - \sin(\phi))$. This convention expresses the idea that, if we push harder in the x -direction, material will move along the y -direction.

Combining these assumptions, we finally have the governing differential equation

$$\frac{\partial}{\partial y}\bar{\sigma} - \frac{2\eta K\bar{\sigma}}{D} = -\rho G. \quad (8)$$

Solving, and using the condition that $\bar{\sigma} \rightarrow 0$ as $y \rightarrow H$ gives

$$\bar{\sigma} = \frac{\rho GD}{2\eta K} \left(1 - \exp\left(\frac{2\eta K}{D}(y - H)\right) \right). \quad (9)$$

The important point to notice is that, as $y \rightarrow 0$, the average stress in a granular material stored in a bin approaches an asymptotic value $\rho GD/2\eta K$, independent of how tall the bin is. This behavior is much different than we found for a liquid-filled bin.

To give you some idea of the numbers involved, here are some typical values for a sand stored in a bin.

- The mass density is about 1.3 g/cm^3 .
- A typical bin diameter might be 5 meters.
- The internal friction angle $\phi \approx 35^\circ$.
- The wall friction angle $\phi_w \approx 15^\circ$.
- The height of fill in a bin is many times the diameter usually; $H = 5D$ is a reasonable guess.

These values give an asymptotic value of the stress of $\bar{\sigma} \approx 400G$. By way of contrast, if the bin had been filled with water (whose density is 1 g/cm^3) to the same height, the pressure at the bottom would be about $10^8 G$. The reason for this large difference is the internal friction. In effect, the internal friction transfers much of the weight of a granular medium to the walls; each bit of the bin wall holds up its share of the material inside. In the case of a fluid like water, all the weight is supported by the material below; thus the bottom of the bin is made to support all the weight, leading to high pressures.

Converging hopper. FIGURE 6 shows the forces in a converging hopper. Mimicing the ideas we just used, we derive the differential equation for the average stress

$$\frac{\partial}{\partial y}(A(y)\bar{\sigma}) - \frac{2\eta K\bar{\sigma}}{\sin(\theta_w)} = -\rho GA(y). \quad (10)$$

The major difference between Equations 8 and 10 is the variable width $A(y)$ in (10), instead of the constant width D in (8). To model the tangential force τ , we have again assumed that the x and y -axes are parallel to the eigendirections. This

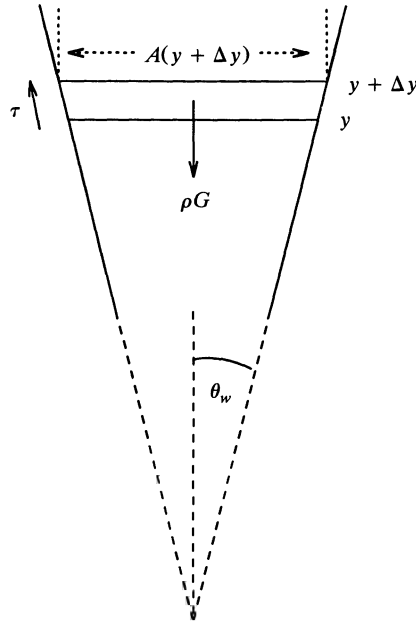


Figure 6. Force balance for a converging hopper.

assumption is not valid for converging bins, but is a reasonable approximation if the walls are steep enough. Using a little trigonometry to find $A(y)$, we have

$$\frac{\partial}{\partial y} \bar{\sigma} + \frac{1}{y} \left(1 - \frac{2\eta K}{\sin(\theta_w)} \right) \bar{\sigma} = -\rho G. \quad (11)$$

Solving out we find

$$\bar{\sigma} = \frac{\sigma_0}{y^{1-\alpha}} - \frac{\rho G y}{2 - \alpha} \quad (12)$$

where $\alpha = (2\eta K/\sin(\theta_w))$. Now we see the assumption of steep walls coming into play. If θ is small enough, $\alpha > 1$ so the first term of Equation 12 looks like $y^{|\alpha-1|}$. That is, both terms decay to zero as $y \rightarrow 0$. Again, this result for granular materials is in sharp contrast to the result for typical fluids. The constant of integration, σ_0 can be used to match the stress at the top of the hopper.

A more precise force diagram for a converging hopper would use polar coordinates with the origin at the (virtual) vertex of the hopper, and stresses σ^{rr} , $\sigma^{r\theta}$, and $\sigma^{\theta\theta}$. The eigendirections approximately align with the radial and circumferential directions; the misalignment is due to the fact that gravity acts vertically. If one carries out such an analysis, one finds $\sigma^{rr} = \rho G f(\theta) r$, where the function $f(\theta)$ is 1 at the centerline of the hopper and decays slowly as $\theta \rightarrow \theta_w$.

DISCUSSION. The previous two sections have introduced the concept of stress and used force diagrams and reasonable simplifying assumptions to describe the pressures that may build up inside a bin containing granular material. We have found that granular material behaves quite differently than typical fluids. The principle reason for this different behavior is the concept of internal friction. In a granular material, each particle resists deformation because of frictional contact with neighboring particles. In a bin, this frictional effect transfers the weight of a slab of material to the wall, reducing the overall pressure inside the bin.

We have not discussed the effect of dynamics in this article. Dynamic effects introduce some additional complexity into the modeling, but some simple problems can be formulated and solved. One such problem is addressed in Savage's paper [2]. The paper by Jenike [1] examines more completely the deformation of granular material inside a hopper in a quasi-static setting.

ACKNOWLEDGMENT. This paper is an expanded version of a talk I gave to the Undergraduate Math Club at UB. I wish to thank the club, and David Jemiolo in particular, for the invitation to speak and for an interesting discussion of granular flow issues.

REFERENCES

1. A. W. Jenike, Steady Gravity Flow of Frictional-Cohesive Solids in Converging Channels, *ASME Journal of Applied Mechanics* **31** (1964) 5–11.
2. S. Savage, The Mass Flow of Granular Materials from Coupled Stress-Velocity Fields, *British Journal of Applied Physics* **16** (1965) 1885–1888.

Department of Mathematics
State University of New York
Buffalo, NY 14214-3093
pitman@galileo.math.buffalo.edu

Orderly Currencies

John Dewey Jones

Despite the proliferation of credit cards, most of us take part in at least one cash transaction a day. In these transactions, we often hand over a sum larger than we owe, expecting to receive the right change. In 1991, the average American spent 12.3 hours making or waiting for change.*

Under these circumstances, it is clearly of the highest importance to develop a robust and efficient change-giving algorithm. An acceptable algorithm must also use the smallest number of bills or coins to give the correct change. Giving change in pennies is mathematically straightforward, but places an unacceptable burden on the nation's pockets and purses.

To develop a *general* algorithm, that is, one that will work in all currency systems, is impossibly difficult. But there is an algorithm that will work in any currency system meeting certain conditions. In this note, we set out the algorithm and establish the necessary and sufficient conditions that the currency system must meet for it to yield the right change. By observing these conditions when devising changes to the currency system, Treasury officials and dictators of newly created states can greatly increase the efficiency of cash transactions within their domains.

Definition. A *currency* is an ordered set of rationals, d_1, \dots, d_n , with $d_1 < d_2 < \dots < d_n$. We refer to the d_i as *denominations*.

Remark. The extension to the rationals is necessary to take into account fractional denominations, such as the half-penny, the farthing, and the groat. The use of fractional denominations, though picturesque, greatly complicates financial transactions, and it is our recommendation that they be eliminated. We can thus restrict the d_i to the integers.

We are not aware of any currency having negative denominations. Without wishing to prejudge the advantages of such a scheme, we shall, in the remainder of this paper, restrict the d_i to the natural numbers.

Definition. A currency is *inflated* if the set d_i has a highest common factor greater than unity.

Remark. Due to inflation or other historical reasons, some currencies have all their denominations being multiples of some integer greater than unity. For example, the base unit of Kenyan currency is the cent, approximate value \$US0.001, but there is no 1-cent coin, only 5-cent and 10-cent coins. In the remainder of this paper, we shall consider that all currencies have been *deflated* by dividing their denominations by their highest common factor.

*All statistical data in this note have been made up in order to support my argument.

Definition. For a given currency, a *payout of value N* is an ordered set b_1, \dots, b_n , where the b_i are non-negative integers and

$$\sum_{i=1}^n b_i d_i = N.$$

Definition. The *bulk* of a payout is the sum $B_N = \sum_{i=1}^n b_i$.

Remark. It might be argued that our aim should be to minimise the total *mass* of change given, rather than the total number of coins. This, however, would make the discussion unduly complicated. Instead, we offer the recommendation that all denominations should have a mass in proportion to their value. If this recommendation is adopted, the mass of a given payout becomes fixed.

Definition. The *Change Algorithm* is a method used to generate a payout, as follows:

```

Target = N
For  $i = n$  to 1
 $b_i = \text{int}(\text{Target}/d_i)$ 
Target = Target -  $b_i * d_i$ 
Next  $i$ 

```

Definition. The *Change Series* is the unique series $b_1 \dots b_n$ generated for a given sum N and a given currency $d_1 \dots d_n$. It can be characterised by the relationships

$$\begin{aligned}
b_n d_n &\leq N < (b_n + 1) d_n \\
b_{n-1} d_{n-1} &\leq N - b_n d_n < (b_{n-1} + 1) d_{n-1} \\
&\vdots \\
b_1 d_1 &= N - \sum_{i=2}^n b_i d_i.
\end{aligned}$$

Remark. The definitions given so far are not adequate to ensure that the Change Algorithm will always work. For example, consider the currency $(2, 3)$, which fails when trying to make a payout of 7. A sufficient condition to ensure that the Change Algorithm does always work is that $d_1 = 1$. In the remainder of the paper, we shall consider only cases where this condition is met.

Definition. A currency is *orderly* if the bulk of the payout generated by the Change Algorithm is less than or equal to the bulk of any other payout of the same value.

Theorem. *A necessary and sufficient condition for a currency to be orderly is that it meets the condition*

$$d_j \geq 2d_{j-1} - d_{j-2}, \quad 3 \leq j \leq n.$$

Proof: Suppose the theorem holds for currencies having up to m denominations. (We note that it holds for a currency having just one denomination.) To prove sufficiency, we start off by supposing $b_{m+1} = \text{int}(N/d_{m+1})$. If this really is the best value for b_{m+1} , we have established the inductive hypothesis. If it isn't, then b_{m+1}

must be smaller than this, since it can't very well be bigger. But if we reduce b_{m+1} by 1, we've got an extra d_{m+1} to take care of, and since $d_{m+1} \geq 2d_m - d_{m-1}$, the best we can hope for is to increase b_m by 1 and reduce b_{m-1} by 1, which doesn't reduce the sum of the b_i . So putting $b_{m+1} = \text{int}(N/d_{m+1})$ must be the right thing to do, confirming the inductive hypothesis.

Suppose on the other hand that $d_{m+1} < 2d_m - d_{m-1}$. Then $d_{m+1} = 2d_m - d_{m-1} - k$ for some positive integer k . We can represent this k as a sum $\sum c_i d_i$. Now every time we reduce b_{m+1} by 1, we increase b_m by 1, reduce b_{m-1} by 1, and reduce each of the b_i by c_i , which obviously leads to a net reduction in bulk and thus refutes the inductive hypothesis. So we may conclude that the condition is both necessary and sufficient.

RELATED WORK. Exercise 2 on p. 191 of *Fundamentals of Computer Algorithms* by Horowitz and Sahni, discusses a similar problem. They call the 'Changegivers Algorithm' the 'greedy solution', observe that it only works for certain currencies, and claim that any currency in which the lowest denomination is 1 and higher denominations are successive powers of any integer will be orderly, in our sense. This is a stronger condition than the one offered here, and while sufficient, is not necessary.

SUGGESTIONS FOR FUTURE RESEARCH. We might consider the case where the d_i are *functions* rather than constants. In countries suffering hyperinflation, for example, it is often necessary to re-issue all bank-notes every few months, adding a zero to the denomination each time. This involves considerable expense, waste of forestry resources, etc. To obviate this, the currency could be issued bearing a simple formula, such as '\$10^(year-1990)', in place of the conventional denomination. Such a currency would appreciate with time, thus encouraging saving.

Alternatively, this formula could be programmed into a microprocessor incorporated in the note. The processor would then automatically update an LCD display on the face of the note.

ACKNOWLEDGMENTS. This research was not supported by grants DT-45567 and DT-45579 from the United States Department of the Treasury.

*School of Engineering Science
Simon Fraser University
Burnaby, British Columbia
V5A 1S6, CANADA
jones@cs.sfu.ca*

An Interior Fixed Point Property of the Disc

Robert F. Brown and Robert E. Greene

1. STIRRING THE COFFEE. At the end of this evening's meal, stir your cup of coffee and watch the motion of the top of the coffee. You will probably notice that there seem to be points where the coffee is not moving. The Brouwer Fixed Point Theorem, implies that such points must always exist. To be precise, suppose you could set the whole top surface of the coffee in motion at once. Then after a moment, every point of the surface would have moved. We could think of this motion as defining a map, that is a continuous function, $f: D \rightarrow D$ of the closed disc (the top surface) to itself. But Brouwer proved that the disc has the *fixed point property*, that is, that every map from D to itself has a fixed point: a point $p \in D$ such that $f(p) = p$. This contradiction shows that there is a point on the surface where the coffee is not moving.

You will probably observe some fixed points, but they may turn out to be difficult to detect. Since Brouwer's Theorem says nothing about the location of the fixed points, the fixed points might lie only on the boundary of the disc. If the way you stir your coffee produces a map with fixed points of this kind, it may be hard to convince you that Brouwer's Theorem is true because the entire (interior) surface of the coffee would be in motion, with fixed points only where the coffee touches the cup.

If a map $f: D \rightarrow D$ moves every point on the boundary of D , which we shall call S , then f must have a fixed point in $\text{int}(D)$, the interior of D , since f must have a fixed point somewhere on D . But when f does have a fixed point on S , then f may or may not have additional fixed points in $\text{int}(D)$. It is natural to ask what one can conclude about the interior fixed points just from knowing the restriction of f to the boundary S , which we denote by $f|_S$. We will be concerned with the case where $f|_S$ takes S to itself. If $f|_S: S \rightarrow S$ has at least one fixed point, then there is a map $G: D \rightarrow D$ with no interior fixed points such that $G|_S = f|_S$, as we will show in the next section. What is surprising is that, even if $f|_S: S \rightarrow S$ is smooth, that is, continuously differentiable, there may not be any smooth map $G: D \rightarrow D$ with $G|_S = f|_S$ that has no interior fixed points. In joint research with Helga Schirmer [1], we investigated when smooth extensions of maps on the boundary of a compact differentiable manifold must have interior fixed points. Our goal in this article is to explain these ideas in a concrete and easily-visualized setting, the 2-dimensional disc.

Most of the things that can be done continuously in topology can be done smoothly also, because of general results about smooth approximation of continuous functions. But there are some exceptions, usually surprising, to that general principle. We will tell you about a new surprising exception.

2. CONTINUOUS EXTENSIONS WITHOUT FIXED POINTS. We will think of the points of the plane as complex numbers $z = x + iy$ and by D we mean the unit disc where $|z| = \sqrt{x^2 + y^2} \leq 1$. When the point is on the boundary S , we will emphasize this by using the corresponding Greek letter ζ instead.

Now we will present the general construction that we promised in the introduction.

Theorem 1. *Given a map $f: S \rightarrow S$ with at least one fixed point, there exists a map $G: D \rightarrow D$ such that $G|_S = f$ and G has no fixed points on $\text{int}(D)$.*

Proof: Choose a fixed point ω of f . For each point $z \in D$ other than ω , let ζ denote the point that the ray from ω through z intersects S (see FIGURE 1). Since z lies on the line segment with end points ω and ζ , we may write

$$z = (1 - r_z)\omega + r_z\zeta$$

for some r_z with $0 \leq r_z \leq 1$. We define $G: D \rightarrow D$ by setting

$$G(z) = (1 - r_z^2)\omega + r_z^2 f(\zeta).$$

Then G has no fixed points on $\text{int}(D)$ because $G(z)$ is proportionately closer to ω on the line segment between ω and $f(\zeta)$ than z was on the original line segment, as we indicate in FIGURE 1.

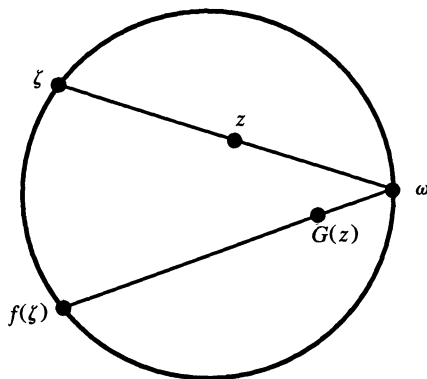


Figure 1

It is not difficult to believe that if f is smooth, the map G we constructed in Theorem 1 is smooth at every point other than ω , but we will next see that, in many cases, no map with the properties of G could be smooth on all of D .

3. THE INTERIOR FIXED POINT PROPERTY. A smooth map of S to itself which does not extend to a map on D with no interior fixed points is the complex squaring map $s(\zeta) = \zeta^2$. If we define $s: D \rightarrow D$ by $s(z) = z^2$, then s has an interior fixed point since $s(0) = 0$, as well as one fixed point $\omega = 1$ on S . You might expect that we could modify s to eliminate the interior fixed point without giving up smoothness. But that is impossible, according to the following result.

Theorem 2 (Interior Fixed Point Property). *Suppose $f: D \rightarrow D$ is a smooth map and such that for some integer $k \geq 2$ we have $f(\zeta) = \zeta^k$ for all $\zeta \in S$, then $f(z) = z$ for some $z \in \text{int}(D)$.*

Thus by the $k = 2$ case of the Property, if we wish to eliminate the interior fixed point of the squaring map $s(z) = z^2$, without changing the map on S , we can hope to do it only if we are willing to give up the smoothness of the map. That is what will happen if we apply the construction of G of the previous section; it gives us a map that fails to be smooth at $\omega = 1$.

If $f: D \rightarrow D$ is a smooth map such that $f(\zeta) = \zeta^k$ for $\zeta \in S$ where $k < 2$, must there still be fixed points in the interior of D ? That is certainly not the case if $k = 0$ since we can just let $f(z) = z^0 = 1$ for all $z \in D$.

We can also find a smooth map f which is the identity on S , that is $f(z) = z^1$, but which has no interior fixed points. Identify the complex number $z = x + iy$ with the pair (x, y) and let

$$f(z) = f(x, y) = \left(x + \frac{1}{2}(1 - x^2 - y^2), y\right).$$

The map f moves all the points in $\text{int}(D)$ by a horizontal motion to the right. It is an elementary exercise to show that the map really takes D to itself, that is, when a point of $\text{int}(D)$ moves to the right it doesn't go outside the unit disc. We can see that f is smooth because its coordinate functions are polynomials. We can also use the same kind of horizontal motion construction for the case $k = -1$. Just set

$$f(z) = f(x, y) = \left(x + \frac{1}{2}(1 - x^2 - y^2), -y\right)$$

then for $\zeta = (x, y) \in S$ we have

$$f(\zeta) = (x, -y) = \zeta^{-1}.$$

When $k \leq -2$, it is still true that the map $f: S \rightarrow S$ defined by $f(\zeta) = \zeta^k$ can be extended to a smooth map $f: D \rightarrow D$ without fixed points on $\text{int}(D)$, as we just did for $k = 1, 0, -1$. However, this is carried out by means of a general construction, rather than by writing down a formula for $f(z)$; see [1], Corollary 7.4.

4. THE FIXED POINT INDEX. Although the Interior Fixed Point Property is a “global” result, a statement about the behavior of a map on all of $\text{int}(D)$, we will see that it depends on the local behavior of a smooth map near its fixed points. The fixed point index, the principal tool of fixed point theory for studying the behavior of a map in a neighborhood of a fixed point, was introduced by Heinz Hopf in the 1920s. We will present Hopf's theory in the setting in which we need it: for maps of subsets of the plane with smooth boundary.

Let M be the closure of an open connected subset of the plane, viewed as the complex numbers \mathbb{C} , whose boundary is a union of smooth simple closed curves. There is a neighborhood U of M such that for each $z \in U - M$ there is a unique point ζ on the boundary of M that is closest to z . (See FIGURE 2.) For the unit disc, we view $M = D$ as the closure of $\text{int}(D)$; then U can be any open disc containing D , and $\zeta = z/|z|$.

Let $f: M \rightarrow M$ be a map whose set of fixed points $\text{Fix}(f) = \{p \in M \mid f(p) = p\}$ is finite. Define $F: U \rightarrow \mathbb{C}$ by

$$F(z) = \begin{cases} z - f(z), & \text{if } z \in M \\ z - f(\zeta), & \text{if } z \in U - M. \end{cases}$$

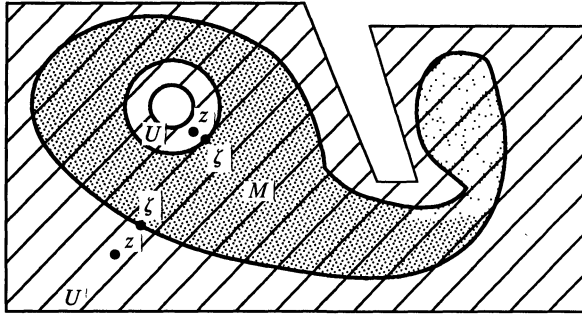


Figure 2

For any $p \in \text{Fix}(f)$ we can find a circle Γ in U about p with the property that p is the only zero of F in the disc bounded by Γ . The *index* $i(f, p)$ of f at p is defined to be the winding number (as in complex analysis) of the closed curve $\gamma = F(\Gamma)$ around the origin 0. We recall that, intuitively, the winding number measures the total number of times γ goes around 0, where counterclockwise rotation counts positively and clockwise negatively. Compare the examples in FIGURE 3. For a rigorous treatment of the winding number, you can refer to, for instance [3], pages 114–116.

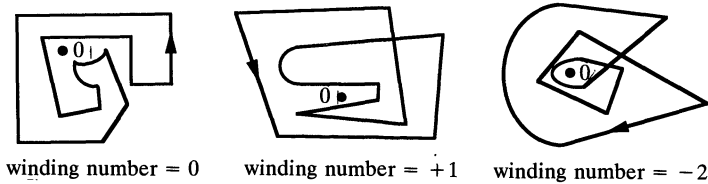


Figure 3

An important property of the winding number is that it is homotopy invariant. In general, a *homotopy* is a map $H: X \times [0, 1] \rightarrow Y$, which may also be thought of as a continuous family of maps $h_t: X \rightarrow Y$ defined by $h_t(x) = H(x, t)$ for $0 \leq t \leq 1$. The map h_0 is said to be *homotopic* to h_1 . In particular, the closed curve $\gamma = F(\Gamma)$ is determined by a map $F|_{\Gamma}: \Gamma \rightarrow \mathbb{C} - 0$ so a homotopy $H: \Gamma \times [0, 1] \rightarrow \mathbb{C} - 0$ defines a family of closed curves $\gamma_t = h_t$. The homotopy invariance of the winding number states that the winding number of γ_0 around the origin equals the winding number of the homotopic curve γ_1 around the origin.

Although the fixed point index is defined locally, Hopf proved an important global result which, in our setting, has the following statement.

Theorem 3 (Hopf [2]). *If M is the closure of an open, connected subset of the plane whose boundary is a union of smooth simple closed curves and $f, g: M \rightarrow M$ are homotopic maps with both $\text{Fix}(f)$ and $\text{Fix}(g)$ finite, then*

$$\sum_{p \in \text{Fix}(f)} i(f, p) = \sum_{q \in \text{Fix}(g)} i(g, q).$$

Hopf's Theorem implies the Brouwer Fixed Point Theorem, as follows. Let $f: D \rightarrow D$ be any map. For the constant map c taking D to the origin, the map $F(z) = z - c(z)$ is the identity map so

$$\sum_{p \in \text{Fix}(c)} i(c, p) = i(c, 0) = 1.$$

Since D is contractible, all self-maps of D are homotopic. If a map f has an infinite number of fixed points there is nothing to prove. If $\text{Fix}(f)$ is finite then, by Hopf's Theorem, when we sum over all fixed points we must always have

$$\sum_{p \in \text{Fix}(f)} i(f, p) = 1,$$

which clearly implies that $\text{Fix}(f)$ is nonempty.

In the next section, we will demonstrate the following.

Theorem 4 (Boundary Fixed Point Index Theorem). *Let $f: D \rightarrow D$ be a smooth map with a finite number of fixed points such that $f(\zeta) = \zeta^k$ for all $\zeta \in S$, for some $k \geq 2$. If π is a fixed point of f that lies in S , then either $i(f, \pi) = 0$ or $i(f, \pi) = -1$.*

The Interior Fixed Point Property follows easily from this result. If $f: D \rightarrow D$ were a smooth map such that $f(\zeta) = \zeta^k$ for all $\zeta \in S$, for some $k \geq 2$ and f had no fixed points on $\text{int}(D)$ then $\text{Fix}(f)$ is finite since $f|_S$ has finitely many fixed points. For each $\pi \in \text{Fix}(f)$ the Boundary Fixed Point Index Theorem tells us that $i(f, \pi)$ is either 0 or -1 so certainly we would have

$$\sum_{p \in \text{Fix}(f)} i(f, p) = \sum_{\pi \in \text{Fix}(f)} i(f, \pi) \leq 0.$$

But that would contradict the consequence of Hopf's Theorem we just used to prove Brouwer's Theorem: if $f: D \rightarrow D$ has a finite number of fixed points, then

$$\sum_{p \in \text{Fix}(f)} i(f, p) = 1.$$

We didn't state the Boundary Fixed Point Index Theorem in its most general form, but rather in a form that made it easy to derive the Interior Fixed Point Property from it. The Index Theorem concerns a local property of the map f , its index at boundary fixed points, and the proof depends only on the smoothness of f near these fixed points, as we will see. Thus f need not be smooth outside of a neighborhood of its fixed point set. A modification of the proof can extend the Boundary Fixed Point Index Theorem to the case where $k \leq 1$, but the conclusion in that case is that either $i(f, \pi) = 0$ or $i(f, \pi) = +1$. That conclusion does not contradict Hopf's Theorem and, as we discussed above, in fact there is a smooth extension of $f(z) = z^k$ without interior fixed points, when $k \leq 1$.

5. PROOF OF THE BOUNDARY FIXED POINT INDEX THEOREM. The fixed point $\pi \in S$ can be written in polar coordinates (r, θ) as $(1, \theta_0)$. If we introduce new coordinates

$$x_1 = \theta - \theta_0 \quad x_2 = 1 - r,$$

then π is the origin 0 in these coordinates and, near π , the boundary S of D corresponds to the x_1 -axis and the interior of D is in the upper half-plane: $x_2 > 0$. Furthermore, since f takes S to itself, f carries this portion of the x_1 -axis to itself.

Near π , the map F has the form

$$F(z) = \begin{cases} z - f(z), & \text{if } z = (x_1, x_2) \text{ with } x_2 \geq 0 \\ z - f(\zeta), & \text{if } z = (x_1, x_2) \text{ with } x_2 < 0, \zeta = (x_1, 0). \end{cases}$$

Writing F in coordinates as $F = (F_1, F_2)$, we observe that (1) $F(0, 0) = (0, 0)$, (2) $F_2(x_1, 0) = 0$, that is, F takes the x_1 -axis to itself, and (3) if $x_2 < 0$, then $F_2(x_1, x_2) < 0$ also.

Conclusion (2) gives us a map g from the x_1 -axis to itself defined by $g(x_1) = F_1(x_1, 0)$. The hypothesis of the Boundary Fixed Point Index Theorem: $f(\zeta) = \zeta^k$ for all $\zeta \in S$ implies that if the map f is restricted to S , then in polar coordinates it has the form $f(1, \theta) = (1, k\theta)$, where, as usual, the polar coordinates angle is measured modulo 2π . Therefore, in the (x_1, x_2) coordinates, $f(x_1, 0) = (kx_1, 0)$ for x_1 near 0. Thus we have

$$g(x_1) = F_1(x_1, 0) = x_1 - kx_1 = (1 - k)x_1.$$

Another hypothesis of the Theorem is that the map $f: D \rightarrow D$ is smooth, that is, has continuous partial derivatives. As we mentioned above, we do not really require the smoothness of f on all of D , but we do need it in a neighborhood of the fixed point, which is the origin in the new coordinates. The smoothness of f does not make the map F smooth in a neighborhood of the origin, since it is not smooth at the x_1 -axis. However, if we let F^+ be the restriction of F to the points (x_1, x_2) near the origin in which $x_2 \geq 0$, then F^+ is a smooth map. Since we are assuming that $k \geq 2$, then

$$\frac{d}{dx_1} F_1(x_1, 0) = g'(x_1) = 1 - k < 0$$

and, therefore, the smoothness of F^+ implies that

$$\frac{\partial F_1^+(x_1, x_2)}{\partial x_1} < 0$$

for (x_1, x_2) in an ϵ -neighborhood of the origin, for $\epsilon > 0$ sufficiently small, and $x_2 \geq 0$. Let Γ be a circle of a radius $\epsilon/2$ about the origin. The negative value of the partial derivative is what we will need to show that the winding number of $\gamma = F(\Gamma)$ about 0 is either 0 or -1 .

Let Γ^+ and Γ^- denote the half-circles above and below the x_1 -axis, respectively. (See FIGURE 4.) By property (3), we know that F maps Γ^- to the “lower” half-plane where $x_2 \leq 0$. To calculate the fixed point index, that is, the winding

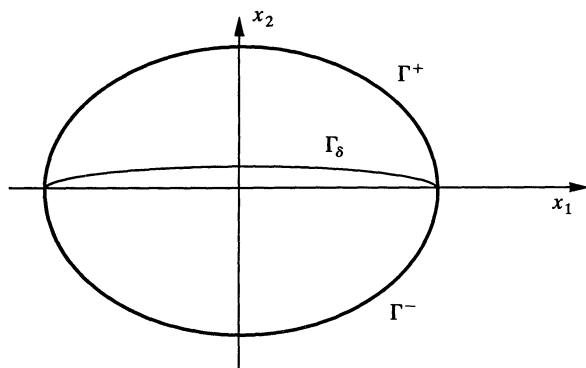


Figure 4

number of $\gamma = F(\Gamma)$ around 0, we need to understand how F behaves on Γ^+ . Since $x_2 \geq 0$ for $(x_1, x_2) \in \Gamma^+$ the map we are concerned with is F^+ . We are assuming that f has only a finite number of fixed points, so we can make ϵ small enough so that the only point on the circle Γ or its interior that F maps to the origin is the origin itself. Therefore, we can homotop the restriction of F^+ to Γ^+ in $\mathbb{C} - 0$ to the restriction of F^+ to the curve Γ_δ for $\delta > 0$, given by

$$\Gamma_\delta^+(t) = \left(\frac{\epsilon}{2}(2t - 1), \delta(1 - (1 - 2t)^2) \right),$$

where $0 \leq t \leq 1$. See FIGURE 4, and notice that for δ small, Γ_δ^+ is quite flat; it *almost* just runs along the x_1 -axis from $(-(\epsilon/2), 0)$ to $(\epsilon/2, 0)$. The homotopy avoids 0 so, by homotopy invariance, it will not change the winding number.

We write the restriction of F^+ to Γ_δ^+ in coordinates as

$$F^+(\Gamma_\delta^+(t)) = (F_1^+(\Gamma_\delta^+(t)), F_2^+(\Gamma_\delta^+(t)) = (\phi_\delta(t), \psi_\delta(t)).$$

We claim that for δ sufficiently small,

$$\frac{d}{dt}\phi_\delta(t) < 0$$

for all t . The intuition is that for δ small, the map ϕ_δ , the restriction of F_1^+ to the “flat” arc Γ_δ^+ is very close to g , the restriction of F_1^+ to the x_1 -axis and since we calculated that $g'(x_1) = 1 - k < 0$, the derivative of ϕ_δ should also be negative since F^+ is smooth: its partial derivatives cannot exhibit abrupt changes. A rigorous proof that the derivative of ϕ_δ is negative can be obtained by a careful application of the chain rule. We furnish this proof as an Appendix to the paper.

Once we accept the claim that ϕ_δ has a strictly negative derivative, what this tells us about the curve $F^+(\Gamma_\delta^+(t)) = (\phi_\delta(t), \psi_\delta(t))$ is that the x_1 -coordinate is a strictly monotone function of t , as we indicate in FIGURE 5. So, in particular, the

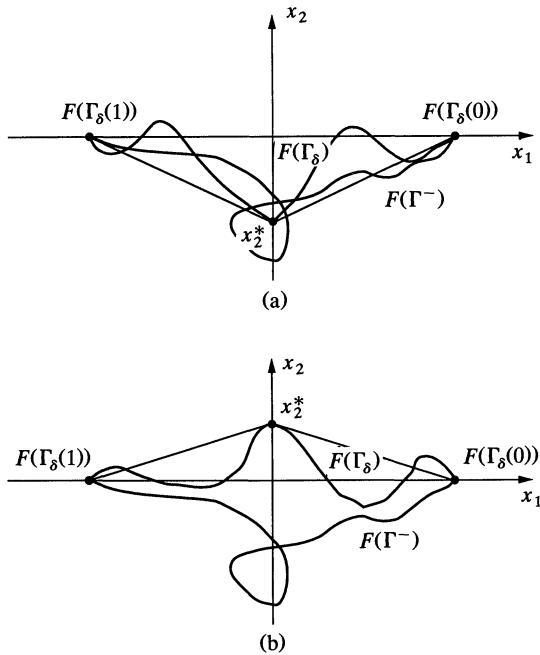


Figure 5

curve intersects the x_2 -axis at just one value, that we denote by x_2^* . As in FIGURE 5, we can homotop the image of $F(\Gamma_\delta(t))$ to the union of line segments

$$[F(\Gamma_\delta(1)), (0, x_2^*)] \cup [(0, x_2^*), F(\Gamma_\delta(0))]$$

by vertical moves, and thus the homotopy avoids the origin. If, as in FIGURE 5a, $x_2^* < 0$, then F takes $\Gamma^- \cup \Gamma_\delta$ entirely into the lower half-plane and the winding number is clearly zero. If $x_2^* > 0$ as in FIGURE 5b, then the winding number of F on $\Gamma^- \cup \Gamma_\delta$ and thus on Γ , is -1 , which completes the proof of the theorem.

6. BEYOND THE DISC. The Boundary Fixed Point Index Theorem is only concerned with the behavior of the map near its boundary fixed points, yet it implies the Interior Fixed Point Property, a statement about the behavior of a map throughout the interior of D . The fact about the disc that we depended on in Section 4 to prove the Property was the contractibility of the disc, which makes all its self-maps homotopic. The disc is the only surface (with or without boundary) that is contractible. (See [4].) Furthermore, the disc and the projective plane are the only surfaces with the fixed point property; on any other surface there is a map that has no fixed points. Thus, in general, we would not necessarily expect other surfaces to share properties of the disc.

If M is a surface with boundary, its boundary consists of circles and we can define a self-map f_C of each boundary component C using the map of S defined by $f(\zeta) = \zeta^k$, by letting $h: C \rightarrow S$ be a homeomorphism and setting $f_C = h^{-1}fh$. An Interior Fixed Point Property on M would be the property that a smooth self-map of M that has some combination of boundary maps of this type must have fixed points in the interior of M . But we don't know of any surface other than the disc that has such a property. For instance, suppose M is the annulus, which we can view as complex numbers z such that $1 \leq |z| \leq 2$ with boundary components S , the unit circle, and $2S$, the circle of radius 2. It is convenient to think of a point of the annulus as $z = t\zeta$ where $\zeta \in S$ and $1 \leq t \leq 2$. Self-maps of the boundary components that are restrictions of a self-map of the annulus are homotopic and therefore a classical result requires them to be of the same degree which, in our terms, means that we must use the same k in $f(\zeta) = \zeta^k$ for each of f_S and f_{2S} . A smooth map that extends these maps of the boundary and has no fixed points on the interior of the annulus can be defined by sending $t\zeta$ to $(1 + (t - 1)^2)\zeta^k$, so the annulus has no Interior Fixed Point Property.

On the other hand, when the Boundary Fixed Point Index Theorem is generalized to manifolds of dimension greater than 2, there are further global consequences. An interested reader can learn more about this from [1] and [5].

APPENDIX. We will prove that

$$\frac{d}{dt}\phi_\delta(t) < 0.$$

Recalling that $\phi_\delta(t) = F_1^+(\Gamma_\delta^+(t))$, where

$$\Gamma_\delta^+(t) = \left(\frac{\epsilon}{2}(2t - 1), \delta(1 - (1 - 2t)^2) \right)$$

we apply the chain rule:

$$\begin{aligned}\frac{d}{dt}\phi_\delta(t) &= \frac{\partial F_1^+}{\partial x_1} \Big|_{\Gamma_\delta(t)} \frac{d}{dt} \left[\frac{\epsilon}{2}(2t-1) \right] + \frac{\partial F_1^+}{\partial x_2} \Big|_{\Gamma_\delta(t)} \frac{d}{dt} [\delta(1 - (1-2t)^2)] \\ &= \frac{\partial F_1^+}{\partial x_1} \Big|_{\Gamma_\delta(t)} \epsilon + \delta \left[\frac{\partial F_1^+}{\partial x_2} \Big|_{\Gamma_\delta(t)} (4-8t) \right].\end{aligned}$$

We can choose $\eta > 0$ small enough so that the closed bounded set

$$\{\Gamma_\delta(t) : 0 \leq \delta \leq \eta, 0 \leq t \leq 1\}$$

lies within the neighborhood of radius ϵ about the origin, then since we proved in Section 5 that the continuous function $\partial F_1^+/\partial x_1$ is negative on that neighborhood, in the upper half-plane, it follows that there exists $A < 0$ so that

$$\frac{\partial F_1^+}{\partial x_1} \Big|_{\Gamma_\delta(t)} \epsilon \leq A$$

for all $0 \leq \delta \leq \eta, 0 \leq t \leq 1$. On the other hand, since $\partial F_1^+/\partial x_2$ is also continuous, there exists $B > 0$ such that

$$\left| \frac{\partial F_1^+}{\partial x_2} \Big|_{\Gamma_\delta(t)} \right| \leq B$$

for all $0 \leq \delta \leq \eta, 0 \leq t \leq 1$. Therefore, by choosing δ small enough so that

$$\delta \left| \frac{\partial F_1^+}{\partial x_2} \Big|_{\Gamma_\delta(t)} \right| |4-8t| \leq 4\delta B \leq \frac{|A|}{2}$$

we have shown that

$$\frac{d}{dt}\phi_\delta(t) \leq \frac{A}{2} < 0$$

for that choice of δ and all t , as we claimed.

REFERENCES

1. R. Brown, R. Greene and H. Schirmer, Fixed points of map extensions, *Fixed Point Theory and Applications* (Proceedings, Tianjin 1988), vol. 1411, Springer Lecture Notes in Mathematics, 1989, pp. 24–45.
2. H. Hopf, *Über die algebraische Anzahl von Fixpunkten*, Math. Z. 29 (1929), 493–524.
3. N. Levinson and R. Redheffer, *Complex Analysis*, Holden-Day, 1970.
4. W. Massey, *Algebraic Topology: An Introduction*, Harcourt, Brace and World, 1967.
5. H. Schirmer, *Nielsen theory of transversal fixed point sets*, Fund. Math. 141 (1992), 31–59.

Department of Mathematics
University of California
Los Angeles, CA 90024
rfb@math.ucla.edu
greene@math.ucla.edu

Le Cam's Inequality and Poisson Approximations

J. Michael Steele

1. INTRODUCTION. For the sum S_n of n independent, non-identically distributed Bernoulli random variables X_i with $P(X_i = 1) = p_i$, Le Cam [20] established the remarkable inequality.

$$\sum_{k=0}^{\infty} |P(S_n = k) - e^{-\lambda} \lambda^k / k!| < 2 \sum_{i=1}^n p_i^2, \quad (1.1)$$

where $\lambda = p_1 + p_2 + \cdots + p_n$.

Naturally, this inequality contains the classical Poisson limit law (just set $p_i = \lambda/n$ and note that the right side simplifies to $2\lambda^2/n$), but it also achieves a great deal more. In particular, Le Cam's inequality identifies the sum of the squares of the p_i as a quantity governing the quality of the Poisson approximation.

Le Cam's inequality also seems to be one of those facts that repeatedly calls to be proved—and improved. Almost before the ink was dry on Le Cam's 1960 paper, an elementary proof was given by Hodges and Le Cam [18]. This proof was followed by numerous generalizations and refinements including contributions by Kerstan [19], Franken [15], Vervatt [30], Galambos [17], Freedman [16], Serfling [24], and Chen [11, 12]. In fact, for raw simplicity it is hard to find a better proof of Le Cam's inequality than that given in the survey of Serfling [25].

One purpose of this note is to provide a proof of Le Cam's inequality using some basic facts from matrix analysis. This proof is simple, but simplicity is not its *raison d'être*. It also serves as a concrete introduction to the semi-group method for approximation of probability distributions. This method was used in Le Cam [20], and it has been used again most recently by Deheuvels and Pfeifer [13] to provide impressively precise results.

The semi-group method is elegant and powerful, but it faces tough competition, especially from the coupling method and the Chen-Stein method. The literature of these methods is reviewed, and it is shown how they also lead to proofs of Le Cam's inequality.

2. MATRIX PROOF OF LE CAM'S INEQUALITY. If one is charged with the task of producing matrices that might help in understanding the distribution of the sum of n independent non-identically distributed Bernoulli random variables, a little time and thought is likely to lead to n matrices P_i like the $N \times N$ matrix

$$P_i = \begin{pmatrix} 1 - p_i & p_i & 0 & \cdots & 0 & 0 \\ 0 & 1 - p_i & p_i & \cdots & 0 & 0 \\ 0 & 0 & 1 - p_i & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 - p_i & p_i \\ 0 & 0 & 0 & \cdots & 0 & 1 - p_i \end{pmatrix}. \quad (2.1)$$

This is almost a Markov transition matrix, except of course the last row of P_i does not sum to 1. A benefit of this choice of the P_i is that they can be written as

$$P_i = (1 - p_i)I + p_i R,$$

where I is the $N \times N$ identity matrix and R is the $N \times N$ matrix with 1's on the first superdiagonal and 0's elsewhere. Since each of the P_i 's is just a linear combination of I and R , any pair of the P_i commute, and because the matrices P_i are so much like Markov transition matrices, their analysis is still reminiscent of the elementary theory of Markov chains.

In fact, by the usual considerations that attend the multiplication of Markov matrices, you can quickly convince yourself that for $n < N$ the top row of the n -fold matrix product $P_1 P_2 P_3 \cdots P_n$ is given by $(P(S_n = 0), P(S_n = 1), P(S_n = 2), \dots, P(S_n = n), 0, 0, \dots, 0)$, i.e. the first $n + 1$ elements of the top row of $P_1 P_2 \cdots P_n$ correspond precisely to the Bernoulli sum probabilities that we wish to estimate. Also at this point, it may be good to be reminded that N is arbitrary except for the constraint $n < N$, so the padded 0's can go on as far as we like.

So far, we have found a matrix that helps us understand the Bernoulli sum probabilities $P(S_n = k)$, and now we would like to find a matrix that is intimately connected with the Poisson distribution. Given some past experience with calculating matrix functions using the Jordan normal form, one can easily find candidates, but knowledge of Jordan forms is not required. One just needs to compute the exponential of P_i , or, better yet, compute the exponential of a simpler matrix closely connected with P_i .

When we write $P_i = I + Q_i$, we see Q_i has the pleasing form,

$$\begin{pmatrix} -p_i & p_i & 0 & \cdots & 0 & 0 \\ 0 & -p_i & p_i & \cdots & 0 & 0 \\ 0 & 0 & -p_i & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -p_i & p_i \\ 0 & 0 & 0 & \cdots & 0 & -p_i \end{pmatrix} = -p_i I + p_i R, \quad (2.2)$$

and the Poisson distribution emerges clearly when we compute $\exp Q_i$:

$$\begin{aligned} \sum_{k=0}^{\infty} Q_i^k / k! &= \begin{pmatrix} e^{-p_i} & p_i e^{-p_i} & \cdots & e^{-p_i} p_i^{N-1} / (N-1)! \\ 0 & e^{-p_i} & p_i e^{-p_i} & \cdots \\ 0 & 0 & e^{-p_i} & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & e^{-p_i} & p_i e^{-p_i} \\ 0 & 0 & 0 & \cdots & 0 & e^{-p_i} \end{pmatrix} \\ &= \sum_{r=0}^{N-1} p_i^r e^{-p_i} R^r / r!. \end{aligned} \quad (2.3)$$

Note that the Q_i commute, so $\exp(Q_i)\exp(Q_j) = \exp(Q_i + Q_j)$, and, in detail,

$$\prod_{i=1}^n \exp(Q_i) = \exp\left(\sum_{i=1}^n Q_i\right) = \exp(-\lambda I + \lambda R), \quad (2.4)$$

where $\lambda = p_1 + p_2 + \dots + p_n$.

The essence of the method is now fully revealed, and we see the proof of Le Cam's inequality boils down to comparing the top rows of $\prod_{i=1}^n \exp Q_i$ and $\prod_{i=1}^n P_i$. This can be achieved most systematically by introducing matrix norms.

If $A = (a_{ij})$ is any matrix, we set

$$\|A\| = \max_i \sum_j |a_{ij}|. \quad (2.5)$$

This recipe provides a *bona fide* matrix norm, and, in particular, one can easily check the relations $\|AB\| \leq \|A\| \|B\|$, $\|A + B\| \leq \|A\| + \|B\|$, and $\|cA\| = c\|A\|$ for $c > 0$.

It is also easy to use the explicit formulas for P_i , Q_i and $\exp Q_i$ to compute their norms: $\|P_i\| = 1$, $\|Q_i\| = 2p_i$, and $\|\exp Q_i\| \leq 1$. When we compare $\prod P_i$ and $\prod \exp Q_i$ using the norm defined by (2.5) we see that the top row attains the maximum, so we have the basic relation,

$$\sum_{k=1}^{N-1} |P(S_n = k) - e^{-\lambda} \lambda^k / k!| = \left\| \prod_{i=1}^n P_i - \prod_{i=1}^n \exp Q_i \right\|. \quad (2.6)$$

Next, it is easy to check that

$$\begin{aligned} P_1 \cdots P_n - \exp Q_1 \cdots \exp Q_n &= (P_1 - \exp Q_1)(P_2 \cdots P_n) \\ &\quad - (\exp Q_1)(\exp Q_2 \cdots \exp Q_n - P_2 \cdots P_n). \end{aligned} \quad (2.7)$$

This identity virtually completes the proof. We just take norms, use the facts that $\|P_2 \cdots P_n\| \leq 1$ and $\|\exp Q_1\| \leq 1$, then repeat the process on the remaining $(n-1)$ -fold product to obtain

$$\begin{aligned} \|P_1 \cdots P_n - \exp Q_1 \cdots \exp Q_n\| &\leq \|P_1 - \exp Q_1\| \\ &\quad + \|\exp Q_2 \cdots \exp Q_n - P_2 \cdots P_n\| \\ &\leq \sum_{i=1}^n \|P_i - \exp Q_i\|. \end{aligned} \quad (2.8)$$

How should we bound $\|P_i - \exp Q_i\|$? Since $\exp Q_i$ is defined by the expansion for e^x , we naturally look to Taylor's formula, but we should be careful enough to consider a finite expansion with a remainder term. For any smooth function f we have

$$f(1) = f(0) + f'(0) + \int_0^1 (1-u) f''(u) du, \quad (2.9)$$

so, if we let $f(t) = e^{tQ}$, the derivative calculations $f(0) = I$, $f'(0) = Q$ and $f''(u) = Q^2 e^{uQ}$ yield

$$e^Q = I + Q + \int_0^1 (1-u) Q^2 e^{uQ} du. \quad (2.10)$$

Even for functions of matrices, integrals are just the limits of sums, so taking norms inside an integral only makes it larger, i.e. for any $g(u, v)$ we have $\|\int g(u, Q) du\| \leq \int \|g(u, Q)\| du$. Also, just as we computed e^{Q_i} explicitly in order to bound its norm, we can compute e^{uQ_i} explicitly to find $\|e^{uQ_i}\| \leq 1$. Applying these observations to the Taylor representation (2.10), we find

$$\|P_i - e^{Q_i}\| \leq \left\| \int_0^1 (1-u) Q_i^2 e^{uQ_i} du \right\| \leq \frac{1}{2} \|Q_i^2\|. \quad (2.11)$$

This is just the tool needed to bound the right side of (2.8). Stringing together the identity (2.6) with inequalities (2.8) and (2.11), we find

$$\sum_{k=0}^{\infty} |P(S_n = k) - e^{-\lambda} \lambda^k / k!| \leq \frac{1}{2} \sum_{i=1}^n \|Q_i^2\| \leq 2 \sum_{i=1}^n p_i^2, \quad (2.12)$$

where the last inequality depended on $\|Q_i^2\| \leq \|Q_i\|^2$ and our earlier explicit calculation that $\|Q_i\| = 2p_i$. Since $n < N$ is our only restriction on N , we can let $N \rightarrow \infty$ to obtain Le Cam's inequality.

3. THE SEMI-GROUP METHOD HAS VIGOROUS COMPETITORS. Deheuvels and Pfeifer [13] provide a version of Le Cam's inequality that—in an asymptotic sense—has a solid claim on being the last word:

$$\sum_{k=0}^{\infty} |P(S_n = k) - e^{-\lambda} \lambda^k / k!| \sim \sqrt{2/\pi e} \left\{ \sum_{i=1}^n p_i^2 \right\} / \left\{ \sum_{i=1}^n p_i \right\} \quad (3.1)$$

provided $\sum_{i=1}^{\infty} p_i \rightarrow \infty$ and $\max(p_1, p_2, \dots, p_n) \rightarrow 0$ as $n \rightarrow \infty$. The essential ideas behind the proof of (3.1) have been seen in Section 2 in a basic form: one obtains an interpretation of the Bernoulli sum probabilities, introduces a semi-group (like $\exp(tQ)$), finds ways to bound an approximation (like $e^Q - I - Q$), and deals with the difference of two n -fold products. Variations on this pattern are visible in Le Cam [20], Shur [26], and one even can see similar steps in Feller's exposition of Trotter's proof of Lindeberg's central limit theorem.

Through the explicit matrix exponentiation calculations used here, the semi-group method can be seen to be friendly as well as useful. Continued exploration of the method is likely to lead to deep and interesting results, but the semi-group method should not be oversold. There are competitors with considerable power.

The characteristic function method also has a role in Poisson approximations. In particular, the characteristic function method has been used by Rusenko [23] to obtain rates of approximation results for repeated samples taken without replacement, by Presman [22] to obtain refinements of Le Cam's inequality, and by Yakshyavicius [32] to provide an inequality like Le Cam's that is pertinent to classes of discrete distributions other than Bernoulli sums. Despite this litany, applications of the characteristic function method are infrequent in Poisson approximation, and it probably does not rank among the big three: the semi-group method, coupling, and the Chen-Stein method. Enough has been said about the first of these, and it is important to provide some sense of the promise inherent in the other two.

4. THE COUPLING METHOD. Because of its simplicity, the coupling method deserves to be reviewed first. For any two random variables X and Y , we begin by defining their *variation distance* by

$$d(X, Y) = \sup_A |P(X \in A) - P(Y \in A)|. \quad (4.1)$$

For random variables that take values in \mathbb{Z}^+ , the metric $d(\cdot, \cdot)$ has an easily proved alternative expression (cf. Serfling [25], p. 569) that reveals its relevance to Le Cam's inequality:

$$d(X, Y) = \frac{1}{2} \sum_{k=0}^{\infty} |P(X = k) - P(Y = k)|. \quad (4.2)$$

The coupling method is based on the simple observation that for random variables

X and Y defined on the same probability space, one has

$$d(X, Y) \leq P(X \neq Y). \quad (4.3)$$

A second observation that helps the coupling method work well with sums is that for $S_n = \sum_{i=1}^n X_i$ and $S_n^* = \sum_{i=1}^n Y_i$ one has

$$d(S_n, S_n^*) \leq \sum_{i=1}^n d(X_i, Y_i). \quad (4.4)$$

From (4.2), (4.3), and (4.4) we see that a good plan for proving Le Cam's inequality consists of building n bivariate couples $Z_i = (X_i, Y_i)$ such that the Z_i are independent, X_i is Bernoulli with parameter p_i , Y_i is Poisson with parameter $\lambda_i = p_i$, and $P(X_i \neq Y_i)$ is as small as possible. This plan has been successfully pursued in Hodges and Le Cam [18], Freedman [16], Serfling [24], Brown [10], Ahmad [1], and Wang [31]. In fact, the couplings of Serfling, Brown, and Wang all satisfy

$$d(X_i, Y_i) = P(X_i \neq Y_i) = p_i(1 - e^{-p_i}), \quad (4.5)$$

from which Le Cam's inequality (1.1) follows easily. As it happens, there is no difficulty in constructing variables that satisfy (4.5)—just think how to simulate X_i and Y_i simultaneously using a single uniformly distributed random variable.

5. THE CHEN-STEIN METHOD. The Chen-Stein method may be the most powerful method for obtaining Poisson approximations, and it is often as easy to use as the coupling or semi-group methods, even though it may be more subtle conceptually. If one does not stop for motivation, one can say that the Chen-Stein method is based on the fact that for each $\lambda > 0$ and $A \subset \mathbb{Z}^+$ there is a function $x = x_{\lambda, A}: \mathbb{Z}^+ \rightarrow \mathbb{R}$ such that for any non-negative integer-valued random variable T one has the identity:

$$E\{\lambda x(T+1) - Tx(T)\} = P(T \in A) - \sum_{k \in A} e^{-\lambda} \lambda^k / k!. \quad (5.1)$$

Actually, the left-hand side of (5.1) is a natural quantity to consider in the context of Poisson approximation, since by summation by parts one can check that $E\lambda f(T+1) = ETf(T)$ for any f , provided T is Poisson with parameter λ . The identity (5.1) was first developed by Chen [11], and some of its mystery can be removed by studying an analogous identity used by Stein [27] in the context of normal approximations. While it is a good exercise to solve (5.1) for x , all one really needs to know about x is that it is bounded and changes slowly. In particular, Barbour and Eagleson [5] sharpened earlier bounds of Chen [12] and showed that for all A and $\lambda > 0$:

$$|x| \leq \min(1, 4\lambda^{-\frac{1}{2}}) \quad (5.2)$$

and

$$\Delta x = \sup_{m \geq 0} |x(m+1) - x(m)| \leq \lambda^{-1}(1 - e^{-\lambda}). \quad (5.3)$$

From these bounds it is easy to prove—and even sharpen—Le Cam's basic inequality (1.1). If we write $W = S_n$, $W_j = S_n - X_j$, $\lambda = p_1 + p_2 + \cdots + p_n$, and $q_j = 1 - p_j$, we can follow Chen [12] and obtain a second identity that together with (5.1) gives one virtually complete information about the Poisson approxima-

tion. We evaluate the left side of (5.1) as follows:

$$\begin{aligned}
E\{\lambda x(W+1) - Wx(W)\} &= \sum_{j=1}^n E\{p_j x(W+1) - X_j x(W)\} \\
&= \sum_{j=1}^n p_j E\{x(W+1) - x(W_j+1)\} \\
&= \sum_{j=1}^n p_j \{p_j Ex(W_j+2) \\
&\quad + q_j Ex(W_j+1) - Ex(W_j+1)\} \\
&= \sum_{j=1}^n p_j^2 E\{x(W_j+2) - x(W_j+1)\}. \quad (5.4)
\end{aligned}$$

From the Chen-Stein identity (5.1) and the Barbour-Eagleson bound on Δx , we see that (5.4) gives

$$\sup_A \left| P(S_n \in A) - e^{-\lambda} \sum_{k \in A} \lambda^k / k! \right| \leq \lambda^{-1} (1 - e^{-\lambda}) \sum_{j=1}^n p_j^2. \quad (5.5)$$

Since $\lambda^{-1}(1 - e^{-\lambda}) \leq 1$, the identity (4.2) shows that inequality (5.5) is sharper than (1.1). Obviously, the Chen-Stein method is very powerful, though it is only now beginning to be well understood. A richer understanding of the method can be obtained by studying Arratia, Goldstein, and Gordon [2], Barbour [3], Barbour and Eagleson [5, 6, 7], Barbour and Hall [8], Barbour [4], and, of course, Stein [28]. A definitive study of Stein's method and its application to Poisson approximation has recently been given in the volume by Barbour, Holst, and Janson [9].

6. CONCLUSION. Le Cam's inequality provides information on the quality of the Poisson approximation, but it also serves as a talisman that is able to charm concrete insights from general techniques. This survey relied on that second service to illustrate the semi-group method, coupling, and the Chen-Stein method. In the course of these illustrations, it has also been possible to survey most of the work on Poisson approximation since the review of Serfling [25], except for the cascade of work coming from the more refined developments of the Chen-Stein method that are dealt with in detail in the monograph of Barbour, Holst, and Janson [9].

REFERENCES

1. Ahmad, I. A. (1985). On the Poisson approximation of multinomial probabilities, *Statist. Probab. Lett.*, 55–56.
2. Arratia, R., Goldstein, L., and Gordon, L. (1990). Two moments suffice for Poisson approximations: the Chen-Stein method *Statistical Science*, 5, 403–434.
3. Barbour, A. D. (1982). Poisson convergence and random graphs, *Math. Proc. Camb. Phil. Soc.*, 92, 349–359.
4. Barbour, A. D. (1987). Asymptotic expansions in the Poisson limit theorem, *Ann. Probab.*, 15, 748–766.
5. Barbour, A. D. and Eagleson, G. K. (1983). Poisson approximation for some statistics based on exchangeable trials, *Adv. Appl. Probab.*, 15, 585–600.
6. Barbour, A. D. and Eagleson, G. K. (1984). On the rate of Poisson convergence, *Math. Proc. Camb. Phil. Soc.*, 95, 473–480.
7. Barbour, A. D. and Eagleson, G. K. (1986). Random association of symmetric arrays, *Stoch. Anal. Appl.*, 4(3), 239–281.

8. Barbour, A. D. and Hall, P. (1984). On the rate of Poisson convergence, *Math. Proc. Camb. Phil. Soc.*, 95, 473–480.
9. Barbour, A. D., Holst, L., and Janson, S. (1992) *Poisson Approximation*, Oxford University Press, New York, NY.
10. Brown, T. C. (1984). Poisson approximation and the definitions of the Poisson process, *this MONTHLY*, 91, 116–123.
11. Chen, L. H. Y. (1974). On the convergence of Poisson binomial to Poisson distributions, *Ann. Probab.*, 2, 178–180.
12. Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probab.*, 3, 534–545.
13. Deheuvels, P. and Pfeifer, D. (1986). A semigroup approach to Poisson approximation, *Ann. Probab.*, 14, 663–676.
14. Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Volume II, John Wiley and Sons: New York, NY.
15. Franken, P. (1964). Approximation des Verteilungen von Summen unabhängiger nichtnegativer ganzzahliger Zufallsgrößen durch Poissonsche Verteilungen, *Math. Nachr.*, 23, 303–340.
16. Freedman, D. (1974). The Poisson approximation for dependent events, *Ann. Probab.*, 2, 256–269.
17. Galambos, J. (1973). A general Poisson limit theorem of probability theory, *Duke J. Math.*, 40, 581–586.
18. Hodges, S. L. and Le Cam, L. (1960). The Poisson approximation to the binomial distribution, *Ann. Math. Statist.*, 31, 737–740.
19. Kerstan, J. (1964). Verallgemeinerung eines Satzes von Prochorow und Le Cam, *Z. Wahrsch. Verw. Gebiete*, 2, 173–179.
20. Le Cam, L. (1960). An approximation theorem for the Poisson binomial distribution, *Pacific J. Math.*, 10, 1181–1197.
21. Le Cam, L. (1963). On the distribution of sums of independent random variables, *Bernoulli, Bayes, Laplace: Proceeding of an International Research Seminar* (J. Neyman and L. M. Le Cam, eds.), Springer: New York, NY, 179–202.
22. Presman, E. L. (1985). Approximation in variation of the distribution of a sum of independent Bernoulli variables with a Poisson law, *Theory Probab. Appl.*, 30, 417–422.
23. Rusenko, N. (1985). On the approximation of the distribution of some statistics by the Poisson laws, *Lith. Math. J.*, 25(4), 363–370.
24. Serfling, R. J. (1975). A general Poisson approximation theorem. *Ann. Probab.*, 3, 726–731.
25. Serfling, R. J. (1978). Some elementary results on Poisson approximation in a sequence of Bernoulli trials, *SIAM Rev.*, 20, 567–579.
26. Shur, M. G. (1984). The Poisson theorem and Markov chains, *Theory Probab. Appl.*, 29(1), 124–126.
27. Stein, C. (1970). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* Vol. 2, 583–602, University of California Press.
28. Stein, C. (1986). *Approximate Computation of Expectations*, Inst. Math. Statist. Lecture Notes—Monograph Series Vol. 7, Hayward, California.
29. Trotter, H. F. (1959). An elementary proof of the central limit theorem, *Arch. Math.*, 10, 226–234.
30. Vervaat, W. (1969). Upper bounds for the distance in total variation between the binomial and the Poisson distribution, *Statist. Neerlandica*, 23, 79–86.
31. Wang, Y. H. (1986). Coupling methods in Poisson approximation, *Canadian J. Statist.*, 14, 69–74.
32. Yakshyavicius, S. V. (1986). Poisson approximation of the binomial, Pascal, negative binomial and geometric distributions (in Russian), *Litovsk. Mat. Sb.*, 26(1), 165–184.

Department of Statistics
 Wharton School
 University of Pennsylvania
 Philadelphia, PA 19104
 steele@wharton.upenn.edu

Answer to Who Was the Author:
 (p. 14)
 Emmy Noether.

A Generalization of a Theorem of Euler

Dorina Mitrea and Marius Mitrea

One of the most important results of pure incidence in geometry is the so called “two-triangle” theorem discovered by the French architect Girard Desargues (1591–1661) asserting that two triangles are perspective from a point if and only if they are perspective from a line.

Recall that two triangles, with their vertices named in a particular order, are said to be *perspective from a point*, if their three pairs of corresponding vertices are joined by concurrent lines. The intersection point is called *the perspective center* of these triangles (see e.g. [CG]). Alternatively, two triangles are said to be *perspective from a line* if their three pairs of corresponding sides meet in collinear points. The line determined by these points is called *the perspective axis*.

In practice, however, many times we encounter pairs of perspective triangles with supplementary properties, and we shall pay a special attention to those pairs that are also *orthologic*.

In precise terms, we say that the triangles $A_1B_1C_1$ and $A_2B_2C_2$ are orthologic, and we write $\triangle A_1B_1C_1 \sim \triangle A_2B_2C_2$, if the perpendiculars from the vertices A_1, B_1, C_1 on the sides B_2C_2, A_2C_2, A_2B_2 pass through one point, called *the orthology center* of $\triangle A_1B_1C_1$ with respect to $\triangle A_2B_2C_2$.

The symbol $\triangle_1 \sim \triangle_2$ was intended to suggest, and we invite the reader to actually *prove* (by means of the ideas developed in the second part of the paper or otherwise), that orthology is indeed an equivalence relation on the set of all triangles in the Euclidean plane.

Actually, any two arbitrary triangles can be translated and rotated in the plane up to a certain position in which they simultaneously are perspective from a point and orthologic (prove it!), but here are some perhaps more familiar examples of pairs of triangles with these two properties:

- (1) *An arbitrary triangle together with its median triangle* (i.e. the triangle whose vertices are the midpoints of its sides). In this case, the two orthology centers are O and H , the circumcenter and the orthocenter of the larger triangle, respectively. The perspective point is the centroid of the bigger triangle, while the perspective axis is improper.
- (2) *An arbitrary triangle together with its orthic triangle* (i.e. the triangle whose vertices are the feet of the altitudes in the initial one). In this situation, the two orthology centers are H and O , while the perspective center is H . Their perspective axis is called *the orthic axis* of the larger triangle.
- (3) *An arbitrary triangle and the triangle determined by the contact points of its sides with its incircle, or together with the triangle determined by the contact points of its sides with its three excircles, or even with the triangle whose vertices are its three excenters.*

- (4) *An arbitrary triangle together with its interior, or exterior Napoleon triangle (i.e. the triangle determined by the centers of the equilateral triangles erected internally, and externally respectively, on the sides of the initial triangle).*

It is a celebrated result of Leonhard Euler (1707–1783) that in any triangle the orthocenter, the centroid and the circumcenter lie on a line, called *the Euler line* of that triangle. If we want to emphasize the perspective and orthologic relationships, we can restate this result by saying that the orthology centers and the perspective center of an arbitrary triangle with respect to its median triangle lie on a line.

In the light of this observation, the following theorem can be thought of as a generalization of Euler's result.

Theorem. *If two triangles are orthologic and perspective from a point, then the two orthology centers and the perspective center lie on a line which is perpendicular to their perspective axis (if this exists).*

We have already noticed that Euler's configuration occurs precisely for the example (1) above but if we use the second part of our theorem for the example (2) above we obtain even more, namely that *the Euler line of a triangle is perpendicular to its orthic axis*. There are many other interesting applications of this theorem and we invite the interested reader to supply new examples.

The rest of the paper is devoted to proving the main theorem. A basic ingredient in our proof will be the following lemma.

Lemma. $\triangle A_1B_1C_1 \sim \triangle A_2B_2C_2$ if and only if

$$\overrightarrow{MA_1} \cdot \overrightarrow{B_2C_2} + \overrightarrow{MB_1} \cdot \overrightarrow{C_2A_2} + \overrightarrow{MC_1} \cdot \overrightarrow{A_2B_2} = 0, \quad (1)$$

for any point M in the plane.

It is easy to see that the left-hand side of (1) is actually *independent* of M , so that in practice we shall verify (1) for *only one point*. In what follows, we shall use this remark several times.

The proof of this lemma is very simple. On one hand, if the two triangles are orthologic, then (1) is clearly satisfied by choosing M to be the orthology center of $\triangle A_1B_1C_1$ with respect to $\triangle A_2B_2C_2$. On the other hand, if (1) holds for any point in the plane, then for particular choice of M as the intersection point of the perpendiculars dropped from A_1 and B_1 onto B_2C_2 and C_2A_2 , respectively, we get that $\overrightarrow{MC_1} \cdot \overrightarrow{A_2B_2} = 0$. Thus $MC_1 \perp A_2B_2$, so that $\triangle A_1B_1C_1 \sim \triangle A_2B_2C_2$.

Let us digress and indicate how this lemma can be used to prove that e.g. \sim is *reflexive* (this is the only property of \sim we shall need in the sequel). Starting with $\triangle A_1B_1C_1 \sim \triangle A_2B_2C_2$ and taking M to be A_1 in (1) we obtain

$$\overrightarrow{A_1B_1} \cdot \overrightarrow{C_2A_2} + \overrightarrow{A_1C_1} \cdot \overrightarrow{A_2B_2} = 0.$$

But this corresponds to the equality

$$\overrightarrow{MA_2} \cdot \overrightarrow{B_1C_1} + \overrightarrow{MB_2} \cdot \overrightarrow{C_1A_1} + \overrightarrow{MC_2} \cdot \overrightarrow{A_1B_1} = 0$$

written for $M \equiv A_2$, therefore $\triangle A_2B_2C_2 \sim \triangle A_1B_1C_1$.

Consider next two triangles, ABC and EFD , satisfying the hypothesis of the theorem. First we will look at the case when their perspective axis exists. Thus we are assuming that their three pairs of corresponding sides, (AC, DE) , (AB, EF) , (BC, DF) , meet in some points X, Y, Z lying on the perspective axis, which we denote by Δ (see FIGURE 1).

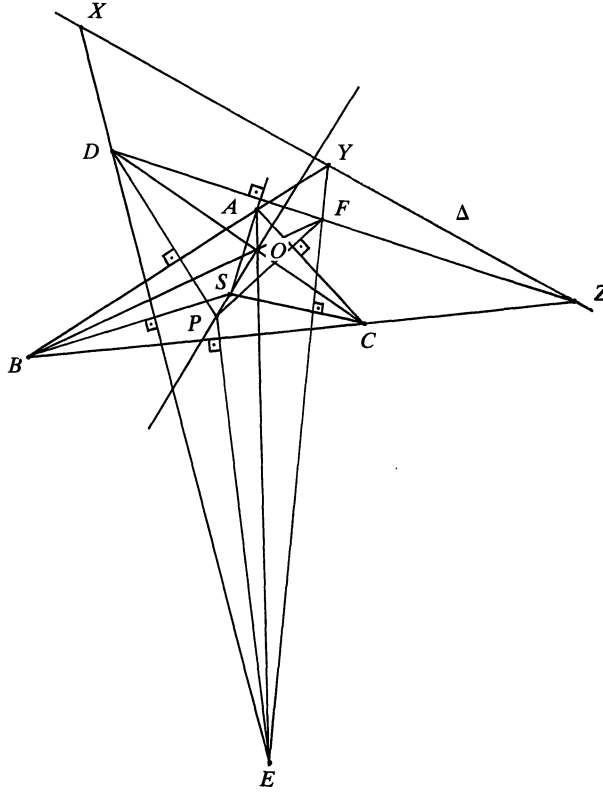


Figure 1

Let O denote their perspective center and let P and S stand for their orthology centers. It suffices to show that $OP \perp \Delta$ since similar reasoning with the roles of S and P interchanged yields $OS \perp \Delta$, i.e. O, S and P lie on the same line.

To this effect, let $\vec{a}, \vec{b}, \vec{c}$ stand for $\overrightarrow{OA}, \overrightarrow{OB}, \overrightarrow{OC}$, so that $\overrightarrow{OD} = \gamma\vec{c}$, $\overrightarrow{OE} = \alpha\vec{a}$, $\overrightarrow{OF} = \beta\vec{b}$, for some real scalars γ, α, β . Clearly, any two of the vectors $\vec{a}, \vec{b}, \vec{c}$ are linearly independent. Furthermore, the assumption that the perspective axis exists ensures that among the real numbers α, β, γ there exists one which is different from the other two. With no loss in generality, let α be that number. Using our Lemma for the triangles ABC and EFG we get

$$\overrightarrow{MA} \cdot \overrightarrow{FD} + \overrightarrow{MB} \cdot \overrightarrow{DE} + \overrightarrow{MC} \cdot \overrightarrow{EF} = 0, \quad (2)$$

for any point M in the plane. In particular, taking $M \equiv O$, (2) becomes

$$\vec{a}(\gamma\vec{c} - \beta\vec{b}) + \vec{b}(\alpha\vec{a} - \gamma\vec{c}) + \vec{c}(\beta\vec{b} - \alpha\vec{a}) = 0. \quad (3)$$

Introducing $\theta := \vec{a} \cdot \vec{b}$, $\psi := \vec{b} \cdot \vec{c}$ and $\varphi := \vec{a} \cdot \vec{c}$, the identity (3) is readily seen to be equivalent to

$$\alpha(\theta - \varphi) + \beta(\psi - \theta) + \gamma(\varphi - \psi) = 0. \quad (4)$$

Going further, $OP \perp \Delta$ is equivalent to $\Delta DOF \sim \Delta XAY$ or, once again in virtue of our Lemma and the remark immediately following it, to

$$\overrightarrow{MD} \cdot \overrightarrow{AY} + \overrightarrow{MO} \cdot \overrightarrow{YX} + \overrightarrow{MF} \cdot \overrightarrow{XA} = 0$$

for some point M in plane. Choosing $M \equiv O$, we are left with proving that

$$\overrightarrow{OD} \cdot \overrightarrow{AY} + \overrightarrow{OF} \cdot \overrightarrow{XA} = 0. \quad (5)$$

Let $\overrightarrow{OY} = \overrightarrow{OA} + \lambda(\overrightarrow{OA} - \overrightarrow{OB}) = \overrightarrow{OE} + \mu(\overrightarrow{OE} - \overrightarrow{OF})$. As \vec{a} and \vec{b} are linearly independent, it follows that $\lambda = \beta(1 - \alpha)/(\alpha - \beta)$, i.e. $\overrightarrow{AY} = \overrightarrow{OY} - \overrightarrow{OA} = \lambda(\overrightarrow{OA} - \overrightarrow{OB}) = \beta(1 - \alpha)/(\alpha - \beta)(\vec{a} - \vec{b})$.

In a similar fashion we get $\overrightarrow{XA} = \gamma(1 - \alpha)/(\alpha - \gamma)(\vec{c} - \vec{a})$. Consequently, (5) becomes

$$\gamma\vec{c} \frac{\beta(1 - \alpha)}{\alpha - \beta}(\vec{a} - \vec{b}) + \beta\vec{b} \frac{\gamma(1 - \alpha)}{\alpha - \gamma}(\vec{c} - \vec{a}) = 0$$

$$\Leftrightarrow \frac{\varphi - \psi}{\alpha - \beta} + \frac{\psi - \theta}{\alpha - \gamma} = 0$$

$$\Leftrightarrow (\alpha - \gamma)(\varphi - \psi) + (\alpha - \beta)(\psi - \theta) = 0$$

$$\Leftrightarrow \alpha(\varphi - \theta) + \beta(\theta - \psi) + \gamma(\psi - \varphi) = 0,$$

i.e. (5) is equivalent to (4), and the conclusion follows in this case.

Finally, in the case in which we do not have a proper perspective axis (i.e. this is thrown to infinity), we infer that $\alpha = \beta = \gamma$. Hence $\triangle ABC$ and $\triangle EFD$ are homothetic with constant α . Now, the orthology centers of $\triangle EFD$ and $\triangle ABC$ become the orthocenter of $\triangle EFD$ and the orthocenter of $\triangle ABC$, respectively.

Since these triangles are homothetic, it follows that O , S and P lie on the same line. Moreover, $\overrightarrow{OS} = \alpha \overrightarrow{OP}$. This completes the proof of the Theorem.

ACKNOWLEDGMENT. The authors thank the anonymous referee whose suggestions led to this improved version of the original manuscript.

REFERENCE

[CG] Coxeter, H., S., M. and Greitzer, S., L., *Geometry Revisited*, Random House 19 (1967).

*Department of Mathematics
University of South Carolina
Columbia, SC 29208*

The Existence of a Triangle with Prescribed Angle Bisector Lengths

Petru Mironescu and Laurentiu Panaitopol

Given three arbitrary positive numbers m, n, p , does there exist a triangle with angle bisectors of length m, n, p ? The answer is YES! Moreover, the triangle is unique up to an isometry. This contrasts with related results for the triangle. The median lengths m, n, p satisfy $m < n + p$ (and the two other cyclic inequalities) and the altitude lengths satisfy $1/m < 1/n + 1/p$; thus the lengths m, n, p cannot be arbitrary in these cases.

The referee has kindly submitted to us several bibliographic notes which show the long history of this problem. Brocard proposed in 1875 in the *Nouvelle Correspondance Mathématique* the problem of constructing such a triangle. The problem was reduced to that of solving an equation of degree 16 by F. J. Van Den Berg (Nieuw Archief voor Wiskunde, 1889). In 1896, P. Barbarin showed in *Mathesis* that the equation can be chosen to be of degree 14 and that it is irreducible in general. He also showed that this equation becomes an irreducible cubic when two of the angle bisector lengths are equal—which shows the impossibility of a Euclidean construction for the triangle. More detailed references can be found in [1], [2].

We now prove the announced result. First, we derive some needed formulas. If a, b, c are the side lengths, m, n, p the angle bisector lengths and s the semiperimeter of a triangle, then we have

$$m = \frac{2}{b+c} \sqrt{bcs(s-a)} \tag{1}$$

with similar formulas for n and p . We prove (1) by an area argument (see FIGURE

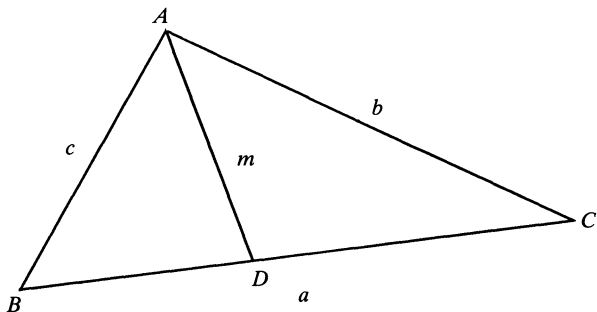


Figure 1

1). If $S(MNP)$ denotes the area of triangle MNP , then

$$2 S(ABC) = 2 S(ABD) + 2 S(ACD)$$

so that

$$bc \sin(A) = bm \sin(A/2) + cm \sin(A/2)$$

and hence

$$m = \frac{bc}{b+c} \frac{\sin(A)}{\sin(A/2)} = \frac{2bc}{b+c} \cos(A/2) = \frac{2bc}{b+c} \sqrt{s(s-a)/bc}.$$

One can easily check that

$$\left[b+c \pm \frac{a(b-c)}{b+c} \right]^2 = 4m^2 + [a \pm (b-c)]^2 \tag{2}$$

and (2) gives

$$b+c = \sqrt{m^2 + (s-b)^2} + \sqrt{m^2 + (s-c)^2}. \tag{3}$$

We get two related equalities for $c + a$ and $a + b$. With the substitutions

$$x = s - a, \quad y = s - b, \quad z = s - c \quad (4)$$

the relation (3) can be rewritten as

$$x = [\sqrt{m^2 + y^2} - y]/2 + [\sqrt{m^2 + z^2} - z]/2 = f(y, m) + f(z, m) \quad (5)$$

where $f(u, v) = [\sqrt{u^2 + v^2} - u]/2$.

Given arbitrary positive real numbers m, n, p , let $F: [0, \infty)^3 \rightarrow (0, \infty)^3$ be defined by

$$F(x, y, z) = (f(y, m) + f(z, m), f(z, n) + f(x, n), f(x, p) + f(y, p)).$$

Taking (5) into account, we get

$$F(x, y, z) = (x, y, z) \quad (6)$$

whenever x, y, z are as in (4) in a triangle with angle bisector lengths m, n, p . Conversely, if (6) holds, then in the triangle with sides lengths $y + z, z + x, x + y$ the equality (3) holds (together with the two analogous ones) so that m must be given by (1), in virtue of the monotonicity of the right side of (3) in the variable m . Hence the given problem is equivalent to the existence and uniqueness of a fixed point for F .

EXISTENCE. Since $f(u, v) \in [0, v/2]$ for nonnegative u, v , it follows that $F(K) \subseteq K$ where $K = [0, m] \times [0, n] \times [0, p]$. Note that $(0, 0, 0)$ is *not* a fixed point of F . Since K is a convex compact set in \mathbb{R}^3 and F is continuous, the existence follows by the Brouwer Fixed Point Theorem (see, for example, [3]).

UNIQUENESS. For $v \neq 0, u \neq t$, and $D = \sqrt{u^2 + v^2} + \sqrt{t^2 + v^2}$, we have

$$2|f(u, v) - f(t, v)| = |u - t|[1 - (u + t)/D] < |u - t|. \quad (7)$$

For $(x, y, z) \neq (x', y', z')$, (7) gives

$$\begin{aligned} |F(x, y, z) - F(x', y', z')| &< (1/2)\sqrt{\Sigma(|y - y'| + |z - z'|)^2} \\ &\leq \|(x, y, z) - (x', y', z')\| \end{aligned} \quad (8)$$

where Σ stands for the cyclic sum and $\| \cdot \|$ denotes the Euclidean norm in \mathbb{R}^3 . Uniqueness follows immediately from (8).

REFERENCES

1. Nathan Altshiller Court, The problem of the three bisectors, *Scripta Mathematica*, XIX (June–September 1953), 218–219.
2. O. Bottema, A theorem of F. J. Van Den Berg (1833–92), *Nieuw Archief voor Wiskunde* (3), XXVI (1978), 161–171.
3. John Milnor, Analytic proofs of the ‘Hairy Ball Theorem’ and the Brouwer Fixed Point Theorem, this MONTHLY, 85 (1978), 521–524.

*Department of Mathematics
University of Bucharest
Bucharest, Romania*

THE EVOLUTION OF . . .

Edited by Abe Shenitzer

Mathematics, York University, North York, Ontario M3J 1P3, Canada

An English major may or may not be a novelist or a poet, but would undoubtedly be expected to be able to evaluate a novel or a poem. The term “English major” implies some historical, philosophical, and evaluative training and competence. It is sad but true that the term “mathematician” does not imply corresponding training and competence.

Integration of the narrowly mathematical and historical, philosophical and critical aspects of our discipline is bound to make it more meaningful not only to those who identify themselves as mathematicians but also to those who have no more than a tangential interest in the subject.

To promote such integration, and thus encourage an approach to mathematics that emphasizes its meaning and significance, the Monthly will publish every two months an article of 2–5 pages under the generic title “The evolution of . . .” The core of such an article will be an account of important mainstream mathematics. The essay that follows exemplifies the kind of material, and the approach, we have in mind.

While we prefer original articles, we will also publish translations or adaptations of appropriate articles in the public domain.

Abe Shenitzer

The Evolution of Integration

A. Shenitzer and J. Steprāns

THE GREEK PERIOD. The Greek problem underlying integration is the *quadrature problem*: Given a plane figure, construct a square of equal area.

It is easy to solve the quadrature problem for a polygon, a figure with rectilinear boundary. The first quadrature of a figure with curvilinear boundary was achieved by Hippocrates in the fifth century B.C. Hippocrates showed that the area of the lunule in FIGURE 1 (that is, the figure bounded by one-half of a circle of radius 1 and one-quarter of a circle of radius $\sqrt{2}$) is equal to the area of the unit square B .

Hippocrates managed to square two other lunules.*

In the third century B.C. Archimedes effected the quadrature of a parabolic segment. He showed that its area is $\frac{4}{3}\Delta$, where Δ is the triangle of maximal area inscribed in the parabolic segment.

Archimedes effected a number of other quadratures (and cubatures). Some of his quadratures involved inventive constructions but most relied on the technique of wedging an area between ever closer upper and lower approximating sums.

*Two more quadrable lunules were found by T. Clausen in the 19th century. In the 20th century, two Russian algebraists proved (independently) that these five lunules are the only quadrable ones.

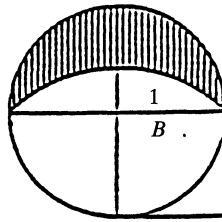


Figure 1

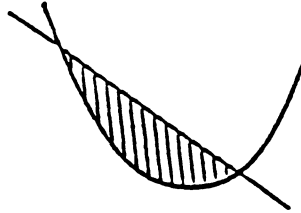


Figure 2

Analogues of such sums are a key element of the definition of the Darboux integral (a variant of the Riemann integral introduced by Darboux in the 19th century) as well as of quadrature programs for computers. We illustrate both of Archimedes' approaches next.

Consider FIGURE 3. Here the hypotenuse AB of the right triangle OAB is tangent to the spiral at A . It then turns out that the side AB is equal to the

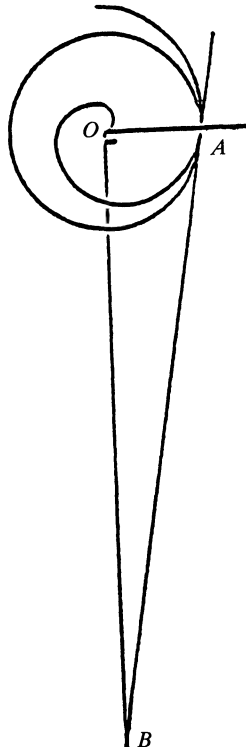


Figure 3

circumference of the circle with radius OA . (This is a special case of Archimedes' rectification of circular arcs by using tangents to spirals.) Since he knew that the area of a circle is half the product of its circumference by its radius, we can say that *Archimedes used (a tangent to) a spiral to rectify a circle and square its area*. Their brilliance notwithstanding, such constructions have been reduced to historical footnotes because they failed to yield general methods.

FIGURE 4 shows a turn of Archimedes' spiral $r = a\theta$ and the associated circle of radius $2\pi a$, and thus of area $K = 4\pi^3 a^2$. To compute the area S of the turn of the spiral in FIGURE 4 Archimedes approximates it from below and above by unions of circular sectors indicated in FIGURE 5.

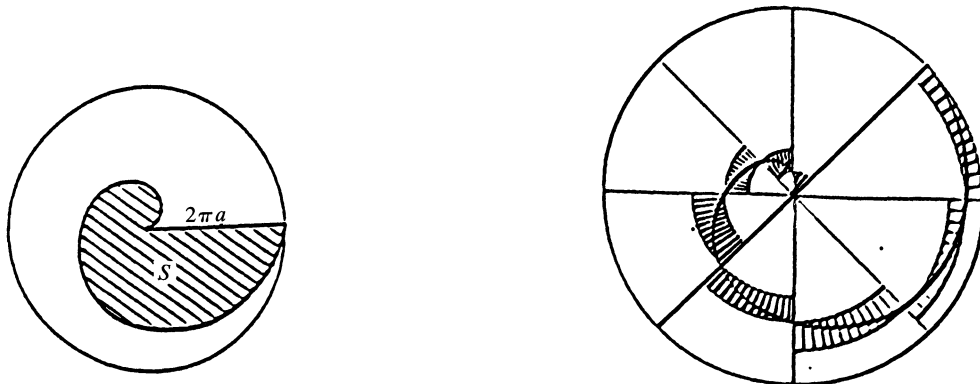


Figure 4 Figure 5

The areas of these approximating figures are, respectively,

$$S'_n = \frac{4\pi^3}{n^3} [1^2 + 2^2 + \cdots + (n-1)^2] = \frac{2\pi^3 a^2 (n-1)(2n-1)}{3n^2}$$

and

$$S''_n = \frac{4\pi^3}{n^3} [1^2 + 2^2 + \cdots + n^2] = \frac{2\pi^3 a^2 (n+1)(2n+1)}{3n^2}.$$

It is not difficult to see that

$$S'_n < \frac{4}{3}\pi^3 a^2 < S''_n$$

for all n . This double inequality can be rewritten as

$$S'_n < \frac{K}{3} < S''_n$$

for all n . Obviously,

$$S'_n < S < S''_n$$

for all n . To prove that $S = (K/3)$ Archimedes shows that $S''_n - S'_n = (4\pi^3 a^2 / 3n^2)$ and is thus small for large n . He can now show that the assumption $S \neq K/3$ leads to a contradiction and can conclude that $S = (K/3)$.

While Archimedes makes no explicit use of limits, he relies on the "method of exhaustion," and, in modern terms, the final part of the argument in a proof involving the method of exhaustion (in the above example it is disproving $S \neq K/3$) amounts to proving the uniqueness of the limit of a Cauchy sequence.

CONTINUATION IN THE 17TH CENTURY. Using nonrigorous infinitesimal techniques (rather than rigorous algebraic methods of the kind used by Archimedes) Cavalieri (1598–1647) managed to compute (what we now write as) $\int_0^1 x^k dx$ for $k = 1, 2, \dots, 9$. His chief difficulty was the evaluation of $1^k + \dots + n^k$. In about 1650 Fermat evaluated $\int_0^a x^{p/q} dx$ by means of a brilliant yet simple computation. Further progress was due to Torricelli, Wallace, and Pascal. In particular, Pascal interpreted Cavalieri’s “sum of lines” (the equivalent of area) as a sum of infinitesimal rectangles.

If we combine Fermat’s result with Cavalieri’s understanding of the linearity of the definite integral (*our* terminology!) then we see that by the middle of the 17th century one could evaluate $\int_a^b P(x) dx$, $P(x)$ a “polynomial” with rational exponents.

In 1647 Gregory St. Vincent made a discovery that linked Napier’s logarithm function and the area under the hyperbola $xy = 1$. This connection is now expressed as $\log_e(x) = \int_1^x (dt/t)$.

Newton and Leibniz invented the calculus and made it into a tool with countless applications but neither gave what we would call a rigorous definition of a definite integral (or saw the need for such a definition). Such concerns became dominant in the 19th century.

FROM CAUCHY TO LEBESGUE. The first rigorous definition of a definite integral was given by Cauchy in the 1820s. Cauchy dealt with continuous functions. In view of the importance of Fourier series whose coefficients are given by integrals it was necessary to define the integral for more general functions. This was first done by Riemann. The limitations of the Riemann integral were remedied at the beginning of the 20th century by Lebesgue. An explanation follows.

With each theory of integration there is associated a theory of measure. Specifically, if f is a function on a set E and $f = f^+ - f^-$ (recall that $f^+(x) = \max\{f(x), 0\}$ and $f^-(x) = \max\{-f(x), 0\}$) then $\int_E f$ is defined as the difference $\int_E f^+ - \int_E f^-$ of the measures $\int_E f^+$ and $\int_E f^-$ of the ordinate sets of the nonnegative functions f^+ and f^- respectively.

The measure underlying the Riemann integral is Jordan measure and the measure underlying the Lebesgue integral is Lebesgue measure. How do they differ? In what way is one “better” than the other?

Consider the simple case of the ordinate set M of a bounded, nonnegative function f on an interval, $0 \leq f(x) \leq c$ for x in $[a, b]$. The Jordan measure of M is the common value, if any, of the outer and inner Jordan measures of M . The outer Jordan measure of M is the glb of the areas of the coverings of M consisting of *finite* unions of rectangles. The inner measure of M is the difference between the area $C(b - a)$ of the rectangle S with base $[a, b]$ and height C and the outer measure of the complement of M in S . Lebesgue replaced the word “finite” in the Jordan definition of the measure of a subset of S by “countable.” This increased greatly the number of measurable subsets of S and led to a theory of integration far more comprehensive and mathematically flexible than Riemann’s.

THE HK-INTEGRAL. Surprisingly, Henstock (in 1955) and Kurzweil (in 1957) came up with a new version of the Riemann integral—call it the HK-integral (see [7])—that is “as good as” the Lebesgue integral! Its definition and main characteristics follow (see [7]):

Definition: A *tagged division* of $[a, b]$ given by a finite ordered set $a = x_0 < x_1 < \dots < x_n = b$ of points, together with a collection of *tags* z_i such that $x_{i-1} \leq z_i \leq$

x_i for $i = 1, \dots, n$. We denote a tagged division by $D(x_i, z_i)$ and the corresponding Riemann sum by

$$S(D(x_i, z_i)) := \sum_{i=1}^n f(z_i)(x_i - x_{i-1}).$$

A *gauge* on $[a, b]$ is a function δ defined on $[a, b]$ such that $\delta(x) > 0$ for all $x \in [a, b]$. An important example of a gauge is a constant function. If δ is any gauge on $[a, b]$, we say that a tagged division $D(x_i, z_i)$ is δ -*fine* in case that $[x_{i-1}, x_i] \subseteq [z_i - \delta(z_i), z_i + \delta(z_i)]$; that is, in case $z_i - \delta(z_i) \leq x_{i-1} \leq z_i \leq x_i \leq z_i + \delta(z_i)$ for all $i = 1, 2, \dots, n$. Finally, we say that the number A is an *HK-integral* of f if, for every $\varepsilon > 0$, there exists a gauge δ_ε such that if $D(x_i, z_i)$ is any tagged division of $[a, b]$ that is δ_ε -fine, then we have

$$|S(D(x_i, z_i)) - A| < \varepsilon.$$

It turns out that “the HK-integral of a function is uniquely defined when it exists and that a function is Riemann integrable if and only if the gauge δ_ε can be chosen to be constant.” More importantly, “every Lebesgue integrable function is HK-integrable with the same value.”

THERE IS NO PERFECT INTEGRAL. While in the eyes of some mathematicians the Lebesgue integral was the final answer to the difficulties associated with integration, there were others who were not willing to give up the search for the *perfect* integral, one which would make all functions integrable. Because Lebesgue’s construction had shown that the key to a comprehensive theory of integration was the construction of an appropriate measure, the search now focussed on finding a total measure on \mathbb{R} , that is, one which assigns a measure to each subset of the real numbers.

Vitali [6] showed that a total measure on the reals cannot be countably additive *and* translation invariant. This being so, it is natural to ask which of these properties should be retained. This decision is, of course, somewhat arbitrary. While retaining translation invariance leads to some fascinating group theory and the Banach-Hausdorff-Tarski Paradox, we will consider what happens if countable additivity is retained instead.

In 1930 S. Ulam [1] showed that there is no such measure on ω_1 , ω_2 or on any cardinal¹ which is the successor of some other cardinal. Ulam’s proof was a spectacular advance in that it did not rely on any of the geometric assumptions, such as translation invariance, on which earlier proofs of the existence of non-measurable sets had relied.

By Ulam’s theorem, the existence of a countably additive measure on \mathbb{R} that measures all of its subsets implies that 2^{\aleph_0} is not the successor of any other cardinal, that is, it is a limit cardinal. By arguing a bit more carefully one can show that there must exist some limit cardinal $\lambda \leq 2^{\aleph_0}$ which is not the union of fewer than λ sets of size less than λ . The existence of such a cardinal has a profound influence on set theory.

In order to understand this influence, it is necessary to recall (a consequence of) Gödel’s second incompleteness theorem which says that set theory can not prove its own consistency. One way to prove the consistency of a theory is to find a model

¹Cantor introduced the notation ω to represent the next ordinal after the integers and it is still favored by set theorists today. The next cardinal after ω is denoted ω_1 and so on.

of that theory, that is, a mathematical structure satisfying all of the axioms of that theory. We ask: What are the implications for set theory of the existence of a model of set theory? Recall the procedure for the construction of the hierarchy of sets. One begins with the empty set—call it V_0 —and then defines V_k to be the power set of V_{k-1} for each integer $k \geq 1$. This is not the end, though, because one can then define V_ω to be the union of the sets V_k and then define $V_{\omega+1}$ to be the power set of V_ω . If one continues this as far as possible and takes the union one gets a model of set theory—or, at least, what would be a model of set theory if it were a set and not a proper class.

How soon, if ever, does this construction process lead to a model of set theory? It turns out that many of the axioms of set theory are satisfied at early stages of the construction. For example the axiom of infinity is satisfied as soon as a single infinite set is included and this is already true of $V_{\omega+1}$. The power set axiom is satisfied at any limit stage because any set which occurs, occurs at a stage before the limit and so all of its subsets are added at the very next stage. The power set itself is therefore added in no more than two stages and, in any case, before the limit. For similar reasons, the pairing axiom is also satisfied at all limit stages. Well-foundedness and comprehension are also easy to deal with.

The problematic axiom is the axiom of replacement, which says that the range of any function defined by a formula is a set. It has already been mentioned that $V_{\omega+\omega}$ will satisfy all of the axioms of set theory except for replacement. Replacement fails because the mapping which takes $2n$ to $\omega + n$ and $2n + 1$ to n is definable by a formula and its domain is ω which belongs to $V_{\omega+1} \subseteq V_{\omega+\omega}$. However, the range of this function is $\omega + \omega$ which does not belong to $V_{\omega+\omega}$. The same argument can be used to show that V_α is a model of set theory if and only if the following holds:

- if $\lambda < \alpha$ then $2^\lambda < \alpha$
- if $\lambda < \alpha$ then any function $F: \lambda \rightarrow \alpha$ (defined using only parameters from V_α) has range bounded in α .

Any cardinal satisfying these requirements is known as a *large* or *inaccessible* cardinal. Since the existence of a large cardinal implies that a model of set theory exists, it follows from Gödel's Theorem that it is impossible to prove the existence of inaccessible cardinals.

Ulam's argument shows that if there is a countably additive measure which measures every set of reals then there is a cardinal α which satisfies the second requirement of being an inaccessible cardinal. Such cardinals are known as weakly inaccessible. Another of Gödel's major contributions is the notion of the Constructible Universe, one of whose consequences is that any model of set theory contains a submodel which satisfies the generalized continuum hypothesis. This allows us to conclude that if there is a weakly inaccessible cardinal then, in the Constructible Universe, the weakly inaccessible cardinal is in fact an inaccessible cardinal; this is so because the cardinal arithmetic of this smaller model of set theory easily implies the first requirement for being a large cardinal.

In other words, if there is a countably additive measure which measures every set of reals then set theory is consistent. This and Gödel's theorem show that the existence of a *perfect* integral is not provable. On the other hand, it is conceivable that some day there may be a proof that it is *not* possible to have a perfect integral. The impact of this on set theory would be devastating. It would follow that many of the large cardinals which experts now consider quite innocuous, and which have played an important role in many important independence results, do

not exist. While this would not show that set theory itself is inconsistent it would severely shake our faith in the assumption that it is.

* * *

We've told our story but would nevertheless like to tack on the following relevant "postscript":

In what sense does the integral solve the Greek quadrature problem and what is its conceptual significance? A telegraphic answer to these two questions follows.

The integral provides a direct "analytic" solution of the Greek quadrature problem for regions of the form

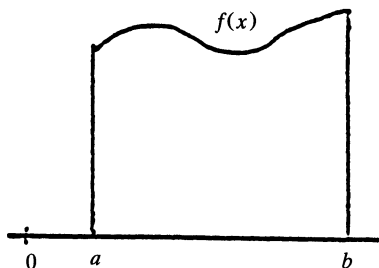


Figure 6

Indeed, the area of the region in the figure is

$$A = \int_a^b f(x) dx.$$

If we rewrite this as

$$A = \int_a^b f(x) dx = (b - a) \left(\frac{1}{b - a} \int_a^b f(x) dx \right),$$

then it is clear that our "integral region" has been replaced by a rectangle of equal area with base $b - a$ and height $(1/(b - a)) \int_a^b f(x) dx$. The quantity $(1/(b - a)) \int_a^b f(x) dx$ is the average of the functional values of f on $[a, b]$. *This averaging ability of the integral is the key to its importance in countless applications.*

REFERENCES

1. S. Ulam, Zur Masstheorie in der allgemeinen Mengenlehre, *Fund. Math.* 16 (1930), 140–150.
2. C. H. Edwards, Jr. *The Historical Development of the Calculus*, Springer-Verlag, 1979.
3. O. Toeplitz, *The calculus—A Genetic Approach*, University of Chicago Press, 1963.
4. A. Aaboe, *Episodes from the Early History of Mathematics*, the MAA, NML 13.
5. T. Jech, *Set Theory*, Academic Press, New York, 1978.
6. G. Vitali, *Sul problema della misura dei gruppi di punti di una retta*, Bologna, 1905.
7. R. G. Bartle, review of R. Henstock's *The General Theory of Integration*, BAMS, v. 29, #1, July 1993, pp. 136–139.

Department of Mathematics and Statistics
York University
North York, Ontario
Canada M3J 1P3.

THE AUTHORS

DAVID A. COX went to Rice University and received his Ph.D. from Princeton University in 1975. After teaching at Haverford and Rutgers, he went to Amherst College in 1979 and has been there ever since. His current area of research is toric varieties, though he has a long-standing interest in arithmetic algebraic geometry and number theory. He is one of the coauthors of the recent undergraduate text *Ideals, Varieties and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*.

THOMAS W. HUNGERFORD received his B.A. from the College of the Holy Cross in 1958 and his Ph.D. from the University of Chicago in 1963, under the direction of Saunders Mac Lane. He served on the faculty of the University of Washington until 1980, when he was appointed Chairman of the Mathematics Department at Cleveland State University, where he is currently Professor of Mathematics. In 1988–89 he was Interim Chairman of the Sociology Department (don't ask). He has written a number of papers in homological algebra and modern algebra texts at both the graduate level (Springer 1974) and undergraduate level (Saunders 1990), as well as several precalculus and business calculus texts.

JOHN STILLWELL received his B.Sc. from Melbourne University, Australia, and his Ph.D. from MIT. Since 1970 he has taught at Monash University and has written several books on mathematics, the most recent being *Mathematics and Its History* (Springer-Verlag 1989) and *Geometry of Surfaces* (Springer-Verlag 1992). He is interested in algebra, geometry, number theory, topology and their history, and is a beginner in Galois theory.

BRUCE PITMAN received a BA in physics from Northwestern University, and a Ph.D. in math from Duke. After post-doctoral appointments at the Courant Institute and the Institute for Mathematics and its Applications at Minnesota, and a year at the New Jersey Institute of Technology, Bruce came to the University at Buffalo in 1989. In addition to his work on granular materials, Bruce has also been involved in a project examining dynamic phenomenon in the kidney. His research incorporates mathematical analysis, modeling, and scientific computing in the study of these problems.

JOHN DEWEY JONES is an Associate Professor at the School of Engineering Science, Simon Fraser University, British Columbia. After graduating from the University of Sussex (UK) with a First in Mathematical Physics, Dr Jones spent several years teaching high school in rural Kenya. He received his Ph.D. in Engineering from the University of Reading (UK) in 1983, and subsequently worked for five years at General Motors Research Laboratories, Michigan. His research interests are in the area of heat transfer, thermodynamics, and the use of artificial intelligence in engineering design.

ROBERT F. BROWN had a relatively late start in mathematics. After an A.B. in Linguistics and English at Harvard, he worked for two years as an operations research analyst. He then went to the University of Wisconsin, where he received a Ph.D. under Edward Fadell. He has been at UCLA since 1963, often teaching courses related to operations research but concentrating his research on topology, especially fixed point theory.

ROBERT E. GREENE received a B.S. from Michigan State University and a Ph.D. from the University of California, Berkeley in 1969. He was a Courant Institute instructor from 1969 to 1971. He has been a member of the UCLA faculty since 1971. His research specialties are differential geometry and several complex variables.

J. MICHAEL STEELE. My interest in probability originated in two wonderful courses that I took from Frank Spitzer during my senior year at Cornell (1970–71). I was fortunately able to go to Stanford to work on a Ph.D., and four years later I completed my dissertation under the direction of Kai Lai Chung. During my first teaching job at the University of British Columbia, I became more involved with statistical issues, and subsequently served as assistant professor of statistics back at Stanford. Fickleness (or breadth of interests) has often seemed to get the best of me, and I eventually served in departments of statistics and computer science (at CMU), statistics, applied and computational mathematics, civil engineering and operations research (at Princeton), and finally back to statistics in the Wharton School of the University of Pennsylvania. I currently serve as editor of the *Annals of Applied Probability*. My interests remain diverse, but the cumulative weight of evidence suggest that the field of probability and algorithms must be where I belong.

WILLIAM MARION received his B.S. degree in mathematics from St. Peter's College in New Jersey in 1965. He completed a Doctor of Arts degree in mathematics at the University of Northern Colorado in 1975 under the direction of Bill Bosch. He has taken graduate level courses in computer science at North Dakota State University, DePaul University and UW-Madison. While on a sabbatical (1991–92) at Madison, he studied theoretical computer science with Eric Bach. Bill came to Valparaiso University in 1978. His interests are in theoretical computer science, discrete mathematics, computer science and mathematics education at the undergraduate level.

IAN STEWART was born in 1945. He obtained a BA at Cambridge University and a Ph.D. at the University of Warwick, where he is now a Professor. He has held visiting positions in Germany, New Zealand, and the USA (Storrs and Houston). His books include *Does God Play Dice?*, *The Problems of Mathematics*, *Another Fine Math You've Got Me Into*, and *Fearful Symmetry: Is God a Geometer?*. He writes the Mathematical Recreations' column in *Scientific American*. His current research field is the effects of symmetry on dynamics, with applications to pattern formation, animal locomotion, and chaos.

JOHN C. POLKING was born on a farm in Iowa. He received his bachelor's degree from Notre Dame and his Ph.D. from Chicago. In between he spent three years in the navy. Following two years as an instructor at Brandeis, he has spent most of the past twenty five years at Rice. Three of those years were spent as Division Director of the Division of Mathematical Sciences at the NSF. His principal research interests include several complex variables and partial differential equations.

Pascal's Matrices

I recently read your article "Pascal's Matrices" in the April, 1993 issue of the *American Mathematical Monthly*. It has some attractive observations about Pascal's triangle, and I applaud your efforts to raise the consciousness of the *Monthly* readership about matrix exponentials. However, I must express sadness that you passed up the occasion to point even higher. Specifically, the theme of your article would have been strengthened, it seems to me, if you had pointed out that *any* one parameter group of matrices *must be* the exponential of some unique matrix, which then may be computed by the method you indicate, of differentiating at the origin. Having said this, you then could have mentioned that this converse fact is basic to the theory of Lie groups. Of the many references to which readers interested in following up these hints might have gone, one of the more convenient for *Monthly* readers might be my article of November 1983.

Roger Howe
Department of Mathematics
Yale University
New Haven, CT 06520

PROBLEMS AND SOLUTIONS

Edited by:

Richard T. Bumby, Fred Kochman and Douglas B. West

Proposed problems should be sent to the MONTHLY PROBLEMS address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.

Solutions of published problems should arrive before June 30, 1994 at the MONTHLY PROBLEMS address given on the inside front cover. Solutions should be typed with double spacing, including the problem number and the solver's name and mailing address. Two copies suffice. A self-addressed postcard or label should be included if an acknowledgment is desired.

*An asterisk (*) after the number of a problem, or part of a problem, indicates that no solution is currently available. Partial solutions will be useful in such cases. Otherwise, the published solution is likely to be based on a solution which is complete and correct. Of course, an elegant partial solution or a method leading to a more general result is always useful and welcome. In addition, references to other appearances of MONTHLY problems or to solutions of these problems in the literature are also solicited.*

10354. *Proposed by Hassan Ali Shāh Ali, Tehran, Iran.*

Determine the least natural number N such that, for all $n \geq N$, there exist natural numbers a, b with $n = \lfloor a\sqrt{2} + b\sqrt{3} \rfloor$.

10355. *Proposed by Joaquín Gómez Rey, I. B. "Luis Buñuel", Alcorcón (Madrid), Spain.*

Two players of equal strength play a tournament consisting of $2n$ matches. Let T be the random variable that counts the number of times the score is tied during the tournament (including the initial 0-0). What is $E(T) + E(T^2)$?

10356. *Proposed by Shalosh B. Ekhad, Princeton, NJ.*

Let X_n be defined by $X_0 = 0$, $X_1 = 1$, $X_2 = 0$, $X_3 = 1$, and for $n \geq 1$,

$$X_{n+3} = \frac{(n^2 + n + 1)(n + 1)}{n} X_{n+2} + (n^2 + n + 1) X_{n+1} - \frac{n + 1}{n} X_n.$$

Prove that X_n is a square of an integer for every $n \geq 0$.

10357. *Proposed by Ira Gessel, Brandeis University, Waltham, MA.*

Define integers, $a_{m,n}$ by

$$\frac{1}{1 - u - v + 2uv} = \sum_{m,n=0}^{\infty} a_{m,n} u^m v^n.$$

Show that $(-1)^j a_{2j, 2j+2}$ is the Catalan number $\binom{2j}{j} / (j + 1)$.

10358*. *Proposed by Jiang Huanxin, student, FuDan University, ShangHai, China*

In triangle $\triangle ABC$, find all points P such that the triangle $\triangle DEF$ (with $D = AP \cap BC$, $E = BP \cap CA$, $F = CP \cap AB$) is equilateral.

10359. *Proposed by Raphael M. Robinson, University of California, Berkeley, CA.*

Two pairs of sides of the unit square $0 \leq x \leq 1$, $0 \leq y \leq 1$ are identified in such a way that the surface obtained has a locally Euclidean metric. How many such surfaces are there which are inequivalent as metric spaces?

10360. *Proposed by Richard P. Stanley, Massachusetts Institute of Technology, Cambridge, MA.*

Let \mathbf{L} be the *integer lattice* in \mathbb{R}^d , i.e. \mathbf{L} is the set of points (x_1, x_2, \dots, x_d) with all $x_j \in \mathbb{Z}$. Consider \mathbf{L} as a graph by declaring two lattice points to be adjacent if the distance between them is 1. Define a sequence S_0, S_1, \dots of subsets of \mathbf{L} inductively as follows:

$$S_0 = \{(0, 0, \dots, 0)\}$$

$$S_n = \left\{ P \in \mathbf{L} - \bigcup_{0 \leq k < n} S_k : P \text{ is adjacent to exactly one element of } \bigcup_{0 \leq k < n} S_k \right\}.$$

Let \mathbf{S} be the subgraph of \mathbf{L} whose vertices are $\bigcup S_n$. Thus, $P \in \mathbf{S}$ is adjacent to $P' \in \mathbf{S}$ if the distance between P and P' is 1.

- Find a simple condition for a point of \mathbf{L} to belong to \mathbf{S} .
- For $P \in \mathbf{S}$, find a simple rule to determine i such that $P \in S_i$.
- How many elements are in S_i ?
- How many $P \in S_i$ are adjacent to no points of S_{i+1} ?
- Show that \mathbf{S} is a tree.
- Investigate the (vertex) density of \mathbf{S} in \mathbf{L} , and compare it to the largest density of a subset of \mathbf{L} for which the induced subgraph is a tree.

NOTES

Notes: (10358) Partial results, giving the properties of such points, would also be of interest. (10360) The case of $d = 2$ already has many features of the general problem. It would be interesting to discover the extent to which this is typical of all dimensions. Note that the only graph structure used in the problem is that induced from the original definition of adjacency in \mathbf{L} .

SOLUTIONS

A Generalized Gamma Function with Independent Branches

6649 [1991, 168]. *Proposed by D. E. Knuth, Stanford University, Stanford, CA.*

Let P be a monic polynomial of degree m with complex coefficients and let ω be a primitive m th root of unity. Let A be the m by m matrix in which the element in the j th row and k th column is

$$A_{jk} = \int_0^\infty (\omega^k t)^j e^{-P(\omega^k t)} dt \quad (0 \leq j, k \leq m-1).$$

Prove that A is nonsingular.

Solution by the proposer. The integrals A_{jk} obviously exist because $P(\omega^k t) = t^m + O(t^{m-1})$ as $t \rightarrow \infty$. Suppose that we have complex numbers $\alpha_0, \dots, \alpha_{m-1}$ such that $\sum_{k=0}^{m-1} \alpha_k A_{jk} = 0$ for $0 \leq j < m$. We wish to prove that $\alpha_0 = \dots = \alpha_{m-1} = 0$. Write $P'(x) = a_0 x^{m-1} + \dots + a_{m-1}$, where $a_0 = m$; then we have, for all $j \geq 0$,

$$\begin{aligned} a_0 A_{j+m,k} + a_1 A_{j+m-1,k} + \dots + a_{m-1} A_{j+1,k} &= \int_0^\infty (\omega^k t)^{j+1} P'(\omega^k t) e^{-P(\omega^k t)} dt \\ &= -\omega^{-k} (\omega^k t)^{j+1} e^{-P(\omega^k t)} \Big|_0^\infty \\ &\quad + (j+1) \int_0^\infty (\omega^k t)^j e^{-P(\omega^k t)} dt \\ &= (j+1) A_{jk}. \end{aligned}$$

Hence, from solving the above for $A_{j+m,k}$, it follows inductively that $\sum_{k=0}^{m-1} \alpha_k A_{jk} = 0$ for all $j \geq 0$. Thus

$$\sum_{k=0}^{m-1} \alpha_k \int_0^\infty Q(\omega^k t) e^{-P(\omega^k t)} dt = 0 \quad (*)$$

for any polynomial Q .

If $P(x) = x^m$, we have $A_{jk} = \omega^{kj} \int_0^\infty t^j e^{-t^m} dt = \omega^{kj} B_j$ where $B_j \neq 0$. Hence $\sum_{k=0}^{m-1} \alpha_k \omega^{kj} = 0$ for $0 \leq j < m$. The Vandermonde matrix (ω^{kj}) is nonsingular, so $\alpha_0 = \dots = \alpha_{m-1} = 0$.

Otherwise, suppose $P(x) = x^m + g(x)$ where $\deg(g) < m$, and let $Q_n(x) = \sum_{k=0}^{n-1} g(x)^k / k!$, so $|e^{g(x)} - Q_n(x)| \leq |g(x)|^n e^{|g(x)|} / n!$. Then, on using $(*)$ with $Q(t) = t^j Q_n(t)$,

$$\begin{aligned} \sum_{k=0}^{m-1} \alpha_k \int_0^\infty (\omega^k t)^j e^{-(\omega^k t)^m} dt &= \sum_{k=0}^{m-1} \alpha_k \int_0^\infty (\omega^k t)^j e^{g(\omega^k t) - P(\omega^k t)} dt \\ &= \sum_{k=0}^{m-1} \alpha_k \int_0^\infty (\omega^k t)^j (e^{g(\omega^k t)} - Q_n(\omega^k t)) e^{-P(\omega^k t)} dt. \end{aligned}$$

Since $g(x) = O(x^{m-1})$ and $|g(\omega^k t)| - P(\omega^k t) = -t^m + O(t^{m-1})$ it follows, for constants K and K_1 independent of n , that

$$\begin{aligned} & \left| \int_0^\infty (\omega^k t)^j (e^{g(\omega^k t)} - Q_n(\omega^k t)) e^{-P(\omega^k t)} dt \right| \\ & \leq \frac{1}{n!} \int_0^\infty t^j |g(\omega^k t)|^n |e^{g(\omega^k t) - P(\omega^k t)}| dt \\ & \leq \frac{K^n}{n!} \int_0^\infty t^{j+n(m-1)} e^{-t^m/2} dt = \frac{2K^n}{m \cdot n!} \\ & \int_0^\infty (2u)^{(j+(n-1)(m-1))/m} e^{-u} du \leq \frac{K_1^n (pn)!}{n!} \end{aligned}$$

where $p = (m - 1/2)/m$ and $j < m$. From Stirling's formula,

$$\lim_{n \rightarrow \infty} K_1^n (pn)! / n! = 0 \quad \text{if } 0 < p < 1,$$

and it follows that $\sum_{k=0}^{m-1} \alpha_k \int_0^\infty (\omega^k t)^j e^{-(\omega^k t)^m} dt = 0$. But this is the case $P(x) = x^m$ already dealt with and so $\alpha_k = 0$ for $0 \leq k < m$.

No other solutions were received.

Tossing Coins Until All Show Heads

E 3436 [1991, 366]. *Proposed by Lennart Råde, University of Göteborg, Sweden.*

Suppose we have n identical coins for each of which heads occurs with probability p . Suppose we first toss all the coins, then toss those which show tails after the first toss, then toss those which show tails after the second toss, and so on until all the coins show heads. Let X be the number of coins involved in the last toss.

- Find $P(X = i)$ for $i = 1, 2, \dots, n$ and $E(X)$.
- Let $p_n = P(X = 1)$. Analyze the behavior of p_n as $n \rightarrow \infty$.

Solution of (a) by Peter Griffin, California State University, Sacramento, CA. The event $X_n = i$ occurs just when, for some k , i coins first show heads on the k th toss while the other $n - i$ coins produced a head at some time before the last toss. Letting $q = 1 - p$, this yields

$$P(X_n = i) = \binom{n}{i} \sum_{k=1}^{\infty} (pq^{k-1})^i (1 - q^{k-1})^{n-i}.$$

(When $i = n$ and $k = 1$, we interpret 0^0 as 1, while for $i < n$ the sum can be started at $k = 2$.) By expanding the binomial, one can write this as a finite sum as follows:

$$P(X_n = i) = \binom{n}{i} p^i \sum_{r=0}^{n-i} \binom{n-i}{r} \frac{(-1)^r}{1 - q^{i+r}}.$$

To calculate $E(X_n)$, recognize that the interior sum of

$$\sum_{i=1}^n iP(X_n = i) = \sum_{k=1}^{\infty} \sum_{i=1}^n i \binom{n}{i} (pq^{k-1})^i (1 - q^{k-1})^{n-i}$$

can be evaluated by regarding it as $(pq^{k-1} + 1 - q^{k-1})^n$ times the mean of a

binomial variable with n trials and success probability $pq^{k-1}/(pq^{k-1} + 1 - q^{k-1}) = pq^{k-1}/(1 - q^k)$, yielding

$$\begin{aligned} EX_n &= \sum_{k=1}^{\infty} n \frac{pq^{k-1}}{1 - q^k} (1 - q^k)^n \\ &= n \sum_{k=1}^{\infty} pq^k (1 - q^k)^{n-1} / q \\ &= \binom{n}{1} \sum_{k=2}^{\infty} pq^{k-1} (1 - q^{k-1})^{n-1} / q \\ &= P(X_n = 1) / q. \end{aligned}$$

Solution of (b) by O. P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands, and the editors. We prove the slightly surprising fact that p_n does not converge as $n \rightarrow \infty$, that instead it oscillates around the pseudo-limit $-p/\log q$ with small, but nevertheless, non-vanishing amplitudes. Let $k_n = \lfloor -\log n / \log q \rfloor$ so that $1 \leq nq^{k_n} < 1/q$. Let $\lambda_n = q^{-\{\log n / \log q\}}$, where $\{x\}$ stands for $x - \lfloor x \rfloor$, the fractional part of x . Then $q^{k_n}/\lambda_n = q^{-\log n / \log q} = 1/n$, which may be written as $nq^{k_n} = \lambda_n$. We may now rewrite p_n as

$$\begin{aligned} p_n &= np \sum_{k=0}^{\infty} q^k (1 - q^k)^{n-1} \\ &= np \sum_{l=-k_n}^{\infty} q^{l+k_n} (1 - q^{1+k_n})^{n-1} \\ &= p \sum_{l=-k_n}^{\infty} \lambda_n q^l \left(\left(1 - \frac{\lambda_n q^l}{n} \right)^{n-1/\lambda_n q^l} \right)^{\lambda_n q^l} \\ &\sim p \sum_{l=-\infty}^{\infty} \lambda_n q^l e^{-\lambda_n q^l}, \end{aligned} \tag{1}$$

where the error in the last approximation approaches 0 as $n \rightarrow \infty$. It is clear that (1) depends on n only through λ_n , which in turn is periodic in the argument $\log n$ (with period $\log q$). Although (1) is reasonably well-approximated by the integral

$$np \int_0^{\infty} q^x (1 - q^x)^{n-1} dx = -p \frac{(1 - q^x)^n}{\log q} \bigg|_0^{\infty} = -\frac{p}{\log q},$$

which depends only on p and q , it is not (quite) independent of n . To see this, write $\lambda_n = q^{-x}$ for $0 \leq x \leq 1$ to form a continuous analog of (1): $f(x) = p \sum_{l=-\infty}^{\infty} q^{l-x} e^{-q^{l-x}}$. In the case of a fair coin ($p = q = 1/2$), this function takes all values between about .721340 and .721355. Formula (1) samples this function at a dense sequence of points; thus p_n will oscillate indefinitely around $1/(2 \log 2) = .7213475 \dots$ without converging.

Editorial comment. Howard Taylor has indicated a method to connect part (b) with more familiar methods of Mathematical Statistics. Each of the n individual coins is replaced by an exponentially distributed random variable η_i with parameter λ , such that $P(\eta_i > x) = e^{-\lambda x}$, for $x > 0$. The number of tosses of a coin is $\lceil \eta_i \rceil$, and X_n is the number of η_i in the interval $(k-1, k]$, where $k = \lceil \max\{\eta_1, \dots, \eta_n\} \rceil$. The analysis leading to (1) is then performed for an exponential distribution.

Bennett Eisenberg, Gilbert Stengle and Gilbert Strang, "The asymptotic probability of a tie for first place", *Annals Appl. Probab.* vol. 3, no. 3 (Aug. 1993) deals with slight variant of this problem. Freed of the *coin-tossing* model, the probability p_j of score j is no longer constrained to follow the geometric distribution. The probability distribution is assumed to be the same for all "players", and the effect of the number of players on the number obtaining the highest score is studied. The results obtained for the geometric distribution are similar to those given above. Slightly more general results were obtained by J. J. A. M. Brands, F. W. Steutel and R. J. G. Wilms, "On the number of maxima of a discrete sample," *Statistics and Probability Letters* (to appear).

Part (a) also solved by R. A. Agnew, M. N. Deshpande (India), O. P. Lossers (The Netherlands), R. M. Robinson, J. Sarkar, P. Warner (student), Western Maryland College Problems group, and the proposer. P. Griffin and the Western Maryland College Problems group also discovered, by computational evidence, the oscillatory behavior of p_n .

Divisibility of the Determinant of a Langley Graph

6657 [1991, 372]. *Proposed by J. J. Rotman and P. M. Weichsel, University of Illinois at Urbana-Champaign.*

Suppose we make a finite group G into a graph Γ by defining two elements a, b of G to be adjacent if $(ab^{-1})^2 \neq e$. If the elements of G are v_1, v_2, \dots, v_n , let A be the n by n matrix in which the element in the i th row and j th column is 1 if v_i and v_j are adjacent and is 0 if v_i and v_j are not adjacent. (A is an adjacency matrix for Γ .)

Show that $\det A$ is an even integer.

Solution I by David Beckwith, Sag Harbor, NY. A determinant is unaltered if any one column is replaced by the sum of all the columns; hence if M is a square matrix with integer entries such that all row-sums are even, then $\det M$ is an even integer.

For the given finite group G , define $X = \{x \in G: x^2 \neq e\}$. Then X has an even number, μ , of elements (possibly zero) because $x \in X$ implies both $x \neq x^{-1}$ and $x^{-1} \in X$. Consider now the i th row of the adjacency matrix A . Its j th entry is 1 if $v_j = x^{-1}v_i$ for some $x \in X$ and 0 otherwise. Thus the row-sum equals μ , independent of i , and the result follows.

Solution II by M. N. Ellingham, Gordon F. Royle and C. C. Timar, Vanderbilt University, Nashville, TN. More generally, the valency of a regular graph is a divisor of the determinant of its adjacency matrix. Let $\phi(M; x)$ denote the characteristic polynomial of an integer matrix M . Then $\phi(M; x)$ is a monic integer polynomial with constant term $\det(M)$. Thus, any rational zero of $\phi(M; x)$ must be an integer dividing $\det(M)$. However, the adjacency matrix of a regular graph of valency k has the column vector $(1, 1, \dots, 1)^T$ as an eigenvector with eigenvalue k .

Now, the argument of Solution I shows that Γ is a regular graph of valency μ , so $\mu | \det A$. Since μ is even, the result follows.

Editorial comment. The two selected solutions were typical of most solutions. In some cases, all algebra with the adjacency matrix was done modulo 2, or modulo μ . Tad White used an argument which paired non-zero terms in the usual expansion of $\det A$ and Paul J. Zweir used a similar pairing on an expression for

the value of $\det A$ in terms of certain subgraphs of Γ (see D. M. Cvetković, M. Doob and H. Sachs, *Spectra of Graphs*, Academic Press, 1979, sect. 1.4, or F. Harary, "The determinant of the adjacency matrix of a graph", *SIAM Review* 4 (1962)).

Thomas Honold noted that the matrix A is $\sum_{g \in X} L(g)$, where L is the left regular representation of G . This matrix is similar to one in block diagonal form with blocks obtained in the same way from the irreducible representations of G . Each irreducible representation occurs with multiplicity equal to its degree, so $\det A$ can be expressed as a product of powers of corresponding determinants for the irreducible representations. The factor arising from the trivial representation is μ , and the factor arising from each other representation is an integer. This is a more elaborate version of the use of eigenvectors in Solution II. However, this approach also identifies other factors of $\det A$. If G contains a central element z of order 2, then any representation restricting to the non-trivial representation of $\{e, z\}$ contributes a factor of zero to $\det A$, so that $\det A = 0$ in this case. Several solvers also noticed this fact, with a direct proof based on noticing that the rows of A corresponding to e and z are the same. On the other hand, it appeared to have escaped notice that the group A_4 also leads to a zero determinant. An infinite family of examples can be generated by considering, for each positive integer k , the group of all linear functions $x \mapsto ax + b$ over a field of $q = 2^k$ elements. This group of order $q(q - 1)$ elements has $q - 1$ one dimensional representations and one $(q - 1)$ dimensional representation. The latter representation is easily seen to contribute a zero factor to $\det A$. Other examples of a similar nature have been found, but a complete characterization of groups G with $\det A = 0$ is unknown, and would be interesting.

Solved also by D. Alvis, D. Astles (U.K.), D. Callan, R. J. Chapman (U.K.), P. Check and D. Birsan (students, Canada), A. J. Douglas (England), W. H. Gustafson, L. Hogben, Th. Honold (Germany), S. Kanetkar, D. W. Koster, C. Lanski, S. C. Locke, O. P. Lossers (The Netherlands), T. White, G. Zappa (Italy), P. J. Zweir, Central Michigan University Problem group, National Security Agency Problems Group, and the proposers.

A Cauchy-Schwarz Inequality for Determinants

E 3464 [1991, 852]. *Proposed by S. J. Bernau and Gavin G. Gregory, University of Texas at El Paso.*

If m and n are given positive integers and if A and B are m by n matrices with real entries, prove that $(\det AB^T)^2 \leq (\det AA^T)(\det BB^T)$.

Composite solution I by many readers. If $m > n$, then AB^T and AA^T are m by m matrices with rank at most n , so $\det(AB^T) = (\det AA^T)(\det BB^T) = 0$, and the desired inequality is obvious, so we may assume $m \leq n$. For any $m \times n$ matrix C and any ascending sequence w of integers $1 \leq w_1 < \cdots < w_m \leq n$, let C_w denote the determinant of the $m \times m$ submatrix of C composed of columns w_1 through w_m . The Binet-Cauchy formula (see F. R. Gantmacher, *The Theory of Matrices*, Chelsea, 1960, vol. I, sect. 1.2) asserts that for any pair A, B of m by n matrices,

$$\det(AB^T) = \sum_w A_w B_w.$$

Here we have summed over all w . Then the assertion of the problem becomes

$$\left(\sum_w A_w B_w \right)^2 \leq \left(\sum_w (A_w)^2 \right) \cdot \left(\sum_w (B_w)^2 \right)$$

which is the Cauchy-Schwarz inequality applied to the $\binom{n}{m}$ -vectors (\dots, A_w, \dots) and (\dots, B_w, \dots) . Thus equality holds iff $A_w = \lambda B_w$ for all w , for some fixed number λ .

The above argument is also valid for complex matrices if transpose is replaced with conjugate transpose.

Solution II by Dennis I. Merino, Southeastern Louisiana University, Hammond, LA. As in solution I, we may assume $m \leq n$. Denoting the i th singular value of a matrix X by $\sigma_i(X)$, we have

$$\begin{aligned} (\det AB^T)^2 &= \prod_{i=1}^m \sigma_i(AB^T)^2 \\ &\leq \prod_{i=1}^m \sigma_i(A)^2 \sigma_i(B^T)^2 \\ &= \prod_{i=1}^m \sigma_i(A)^2 \sigma_i(B)^2 \\ &= \det AA^T \cdot \det BB^T. \end{aligned} \quad (1)$$

The inequality (1) is a special case of Horn's inequality, Theorem 3.3.4 of Roger A. Horn and Charles R. Johnson, *Topics in Matrix Analysis*, Cambridge, 1991.

Solution III by Olaf Krafft, RWTH, Aachen, Germany. Since AA^T and BB^T are nonnegative definite, one has $\det AA^T \geq 0$ and $\det BB^T \geq 0$. Therefore, the inequality is trivially satisfied if $\det AB^T = 0$. Assume that $\det AB^T \neq 0$. Then AB^T is regular, hence $\text{rank } A = \text{rank } B = m$; and, in particular, BB^T is regular. The matrix $I - B^T(BB^T)^{-1}B$ is symmetric and idempotent, thus nonnegative definite. Now, it is a corollary of the Courant-Fischer theorem (see R. Bellman, *Introduction to Matrix Analysis*, McGraw-Hill, 1960 (or second edition, 1970), sect. 7.7) that the ordered eigenvalues of $C + D$ and D satisfy $\lambda_k(C + D) \geq \lambda_k(D)$ if D is symmetric and C nonnegative definite. If D is also nonnegative definite, so that the $\lambda_k(D) \geq 0$, then $\det(C + D) \geq \det D$. Applying this to

$$C = A(I - B^T(BB^T)^{-1}B)A^T, \quad D = AB^T(BB^T)^{-1}BA^T,$$

one gets $\det AA^T \geq \det AB^T(BB^T)^{-1}BA^T = (\det AB^T)^2(\det BB^T)^{-1}$.

Solution IV by Thomas L. Markham, University of South Carolina, Columbia, SC. It is obvious that

$$C = \begin{bmatrix} A \\ B \end{bmatrix} \begin{bmatrix} A^T & B^T \end{bmatrix}$$

is positive semidefinite of order $2m$. If each m by m block of C is replaced by its determinant, the resulting compressed matrix is

$$\begin{bmatrix} \det(AA^T) & \det(AB^T) \\ \det(BA^T) & \det(BB^T) \end{bmatrix}.$$

Using general results about compressed matrices, such as those found in R. C. Thompson, "A determinantal inequality for positive definite matrices", *Canad. Math. Bull.* 4 (1961), 57-62, one finds that this matrix is also positive semidefinite, which proves the result.

Editorial comment. Although the analogous result over \mathbb{C} using the complex conjugate of the transpose was mentioned only briefly in Solution I, several solvers phrased their solutions in those terms, and all approaches to the problem can be modified in this way.

A number of readers gave lengthier solutions along the lines of Solution II, including a proof of the applicable part of Horn's inequality.

Albert Nijenhuis noted that the Binet-Cauchy formula is an explicit coordinate representation for an inner product defined on the m -th exterior power $\wedge^m V$ where V is either \mathbb{R}^n or \mathbb{C}^n . In particular, if $a_1, \dots, a_m, b_1, \dots, b_m \in \mathbb{R}^n$, introduce an m by n matrix A whose i -th row is a_i to describe the element $a_1 \wedge \dots \wedge a_m$ and a matrix B constructed in the same way from the b_i , and define $\langle a_1 \wedge \dots \wedge a_m, b_1 \wedge \dots \wedge b_m \rangle$ to be $\det(AB^T)$. This is shown to be well-defined in texts on Multilinear Algebra (e.g., M. Marcus, *Finite Dimensional Multilinear Algebra*, Part I, Marcel Dekker, 1973). The use of the Cauchy-Schwarz inequality in Solution II leads to $(\det AB^T)^2 = (\det AA^T)(\det BB^T)$ if and only if the exterior product of the rows of A , and that of B are linearly dependent in $\wedge^m V$. When A and B both have rank m , this happens if and only if the two matrices have the same row space.

For further compression inequalities as employed in Solution IV, and generalizations of the problem, Thomas L. Markham supplied the reference: C. R. Johnson and T. L. Markham, "Compression and Hadamard power inequalities", *Linear and Multilinear Algebra* 18 (1985), 23–34; and Roy Mathias supplied references to J. de Pillis, "Transformations on Partitioned Matrices", *Duke Math. J.* 36 (1969), 511–515 and to R. A. Horn and R. Mathias, Cauchy-Schwarz Inequalities Associated with Positive Semidefinite Matrices, *Linear Algebra Appl.* 142 (1990), 63–82.

Solved also by J. M. Arregi and J. Sangoniz (Spain), D. W. Bailey, C. I. Caldwell (student), D. Callan, R. J. Chapman (U.K.), Y. Diao, M. Drešević (Yugoslavia), J. C. Goding, G. Janee, J.-P. Grivaux (France), J. W. Grossman and S. S.-S. Wang, K. P. Hart, L. Hogben, D. Jespersen, P. G. Kirmsner, N. Komanda, R. Mathias, G. Miller (student, Canada), J. M. Monier (France), T. S. Nanjundiah (India), Y. Nievergelt, A. Nijenhuis, I. Olkin, A. Pedersen (Denmark), W. So, F. B. Strauss, A. Tissier (France), G. Trenkler (Germany), M. Vowe (Switzerland), University of Wyoming Problem Circle, and the proposers. Two incorrect solutions were received.

Schröder Numbers Modulo 3

E 3470 [1991, 955]. *Proposed by B.M.M. de Weger, University of Twente, Enschede, The Netherlands.*

Let S_n be the n th Schröder number, defined as the number of polygonal paths in the Cartesian plane that start at $(0, 0)$, end at (n, n) contain no points above the line $y = x$, and are composed of steps taken from the set $\{(0, 1), (1, 0), (1, 1)\}$. E 3343 [1989, 734; 1991, 367] asserted that $S_n \equiv 0 \pmod{3}$ when n is even. Determine S_n modulo 3 when n is odd.

Solution I by Michael Vowe, Therwil, Switzerland. We will show that $S_{2m+1} \equiv (-1)^{m+1} C_m \pmod{3}$ for $m \geq 0$, where $C_n = \binom{2n}{n}/(n+1)$ is the n th Catalan number. In problem E3343, the recurrence for the Schröder numbers: $S_{n+1} = 3S_n + \sum_{k=1}^{n-1} S_k S_{n-k}$ for $n \geq 1$ was obtained, with $S_0 = 1$ and $S_1 = 2$.

Letting $R_m = S_{2m-1} \pmod{3}$, we can reduce the recurrence modulo 3 to obtain $R_{m+1} \equiv \sum_{j=1}^m R_j R_{m+1-j} \pmod{3}$ for $m \geq 1$, with $R_1 = 2$. This uses $S_{2k} \equiv 0 \pmod{3}$.

If we treat this as a recurrence in integers, it will yield numbers congruent to these modulo 3. Let $Q_1 = 2$ and $Q_{m+1} = \sum_{j=1}^m Q_j Q_{m+1-j}$ for $m \geq 1$. Let $f(x) = \sum_{m \geq 1} Q_m x^m$. The recurrence can be expressed as $f(x) - 2x = f^2(x)$. The solution to this is $f(x) = \frac{1}{2}(1 - \sqrt{1 - 8x})$, since $f(0) = Q_0 = 0$. Using the extended binomial theorem, we compute the coefficient of x^m to be $Q_m = (2^m/m) \binom{2m-2}{m-1}$. Since $2 \equiv -1 \pmod{3}$, we obtain $R_{m+1} \equiv (-1)^{m+1} C_m \pmod{3}$, as desired.

Solution II by O.P. Lossers, Eindhoven University of Technology, Eindhoven, The Netherlands. Let $m = m_0 + m_1 \cdot 3 + \dots + m_t \cdot 3^t$ be the ternary expansion of m . Then $S_{2m-1} \equiv \binom{2}{m_0} \binom{1}{m_1} \dots \binom{1}{m_t} \pmod{3}$ for $m \geq 1$. In particular, modulo 3 we have

$$S_{2m-1} \equiv \begin{cases} 0 & \text{if } \lfloor m/3 \rfloor \text{ has a 2 in its ternary expansion,} \\ 2 & \text{if } \lfloor m/3 \rfloor \text{ has no 2 and } m \equiv 1 \pmod{3}, \\ 1 & \text{if } \lfloor m/3 \rfloor \text{ has no 2 and } m \not\equiv 1 \pmod{3}. \end{cases}$$

We begin with the expression $2f(x) = 1 - \sqrt{1 - 8x}$ for $f(x) = \sum Q_m x^m$ found above, which we can rewrite modulo 3 as $f(x) \equiv -1 + (1+x)^{1/2}$. To obtain the desired result, it suffices to show

$$(1+x)^{1/2} \equiv \sum_{m \geq 0} \binom{2}{m_0} \binom{1}{m_1} \binom{1}{m_2} \dots x^m \pmod{3}.$$

Consider the power series $F_n(x) = \sum_{k \geq 0} \binom{(3^n+1)/2}{k} x^k$; we have $[F_n(x)]^2 = (1+x)^{3^n+1}$. Since $(1+x)^{3^n} \equiv 1 + x^{3^n} \pmod{3}$, we have $[F_n(x)]^2 \equiv 1 + x + x^{3^n} + x^{3^n+1} \pmod{3}$. By making n sufficiently large, we can make any desired initial segment of coefficients in $F_n(x)$ agree with those of $(1+x)^{1/2}$ modulo 3.

This yields the desired result, because as n grows, each coefficient $\binom{(3^n+1)/2}{k}$ reaches and remains at the desired value modulo 3. This uses the theorem of Lucas (circa 1878, see L. E. Dickson, *History of the Theory of Numbers*, Chelsea, 1971, vol. I, ch. IX, p. 271), which states that if $a = \sum a_i p^i$ and $b = \sum b_i p^i$, then $\binom{a}{b} \equiv \prod \binom{a_i}{b_i} \pmod{p}$, where p is a prime. Since $(3^n+1)/2 = 2 + 1 \cdot 3 + \dots + 1 \cdot 3^{n-1}$, the coefficient of x^m in $F_n(x)$ reaches and remains at the value described above when $n > \lfloor \log_3 m \rfloor + 1$. Hence Q_m has the value claimed, modulo 3.

Editorial comment. Most solvers gave explicit descriptions of the set of values of n for which S_{2m-1} is congruent to 0, 1, or 2 modulo 3 in terms of the ternary expansion of m . There were various expressions, with the relatively short arguments being those using the formula of Lucas. Fred Galvin and George Kangas extended the results of E 3343 in a different manner by showing for non-negative m that $S_{2m} \equiv 2(-1)^{m+1} \binom{2m}{m} \pmod{9}$. In this work, the more complicated generating function of the Schröder number was used with the congruences modulo 9 entering later in the argument.

Solved also by D. M. Bloom, D. Callan, R. J. Chapman (U.K.), R. High, N. Komanda, A. Nijenhuis, J. H. Steelman, D. M. Wells, Anchorage Math Solutions Group, National Security Agency Problems Group, and the proposer.

Promising Sequences

10185 [1992, 60]. *Proposed by William C. Waterhouse, Pennsylvania State University, University Park, PA.*

If P is a polynomial with coefficients in \mathbb{Z} , it is immediate that $P(b + c) - P(b)$ is an integer multiple of c for all integers b and c . Are there any non-polynomial functions $f: \mathbb{Z} \rightarrow \mathbb{Z}$ with this property?

Solution by Gerry Myerson, Macquarie University, Sydney, NSW, Australia. Yes. Let

$$\begin{aligned} f(n) = & n + n(n^2 - 1^2) + n(n^2 - 1^2)(n^2 - 2^2) \\ & + n(n^2 - 1^2)(n^2 - 2^2)(n^2 - 3^2) + \cdots. \end{aligned}$$

Observe that for each integer n , the sum above consists of only finitely many non-zero terms. We show that f has the desired properties. For each whole number a , let $f_a(n)$ be the sum of the first a terms in the series for f . Then $f_a(n)$ is a polynomial in n with integer coefficients, and $f_a(m) = f(m)$ for all integers m with $|m| \leq a$. Now, given integers b and c , let a be the larger of $|b + c|$ and $|b|$. Then $f(b + c) - f(b) = f_a(b + c) - f_a(b)$ is an integer multiple of c , since f_a is a polynomial with integral coefficients. For positive integers b , the summands in $f(b)$ are all non-negative and the last non-zero summand is $(2b - 1)!$. Thus, $f(b) \geq (2b - 1)!$, whence f grows faster than any polynomial.

Editorial comment. Most of the solutions received were similar to the solution given above. Several solvers pointed out that a more general choice for f along the same lines is

$$f(n) = \sum_{j=0}^{\infty} a_j(n + j)(n + j - 1) \cdots (n - j),$$

where the a_j are integers and infinitely many are non-zero. A few solvers also mentioned polynomials such as $f(n) = (n^2 - n^4)/2$. Note that, while f is a polynomial, it is not in $\mathbb{Z}[x]$, and yet $f(b + c) - f(b)$ is an integer multiple for c for all integers b and c .

A small set of solvers ignored explicit formulas entirely. Starting from any function $n \mapsto a_n$ from \mathbb{N} onto \mathbb{Z} , call a finite sequence of integers b_0, b_1, \dots, b_k *promising* (terminology of Robert High) if it satisfies $(a_i - a_j) | (b_i - b_j)$ for $0 \leq i < j \leq k$. In this approach, one shows that it is always possible to define b_{k+1} in many ways to extend a promising sequence. This demonstrates the existence of uncountably many functions with the required property while there are only countably many polynomials.

Solved also by D. Boixader Ibáñez (Spain), G. Calinescu (student, Romania), R. J. Chapman (U.K.), P. R. Chernoff, F. J. Flanigan, I. Gessel, R. High, K. S. Kedlaya (student), J. H. van Lint (The Netherlands), O. P. Lossers (The Netherlands), G. A. Martin, A. Nijenhuis, J. L. Pietenpol, I. Praton, K. Schilling, N. C. Singer, R. Stong, O. Wyler, Theory First, University of South Alabama Problem Group, and the proposer.

A Family of Invertible Matrices

10214 [1992, 362]. *Proposed by Stephen Penrice, Emory University, Atlanta, GA.*

For all integers $n > 1$, let $f(n)$ denote the largest real number such that, for any set of non-negative real numbers satisfying $a_1 + \cdots + a_n < f(n)$, the n by n matrix with a_1, \dots, a_n along the main diagonal and -1 in all other positions is invertible. Show that $f(n)$ is well-defined, and obtain an explicit formula for it.

Solution by Dirk P. Laurie, Potchefstroom University for Christian Higher Education, Vanderbijlpark, South Africa. The matrix concerned can be written as $\mathbf{M} = \mathbf{D} - \mathbf{e}\mathbf{e}^T$, where $\mathbf{D} = \text{diag}(1 + a_i)$ and \mathbf{e} is a column of ones. By The Sherman-Morrison formula (see Alston S. Householder, *The Theory of Matrices in Numerical Analysis*, Dover, 1964, sect. 5.1),

$$\mathbf{M}^{-1} = \mathbf{D}^{-1}(1 - \mathbf{e}^T \mathbf{D}^{-1} \mathbf{e})^{-1} \mathbf{D}^{-1} \mathbf{e} \mathbf{e}^T \mathbf{D}^{-1};$$

hence \mathbf{M} is invertible whenever $\mathbf{e}^T \mathbf{D}^{-1} \mathbf{e} \neq 1$. Using the fact that the arithmetic mean of positive numbers is not less than their harmonic mean, we obtain

$$\begin{aligned} \mathbf{e}^T \mathbf{D}^{-1} \mathbf{e} &= \sum_{i=1}^n \frac{1}{1 + a_i} \\ &\geq n \left(\sum_{i=1}^n \frac{1 + a_i}{n} \right)^{-1} \\ &= \frac{n^2}{n + f(n)} \\ &> 1 \text{ if } f(n) < n^2 - n. \end{aligned}$$

The function $f(n) = n^2 - n$ is best possible, since if all $a_i = n - 1$, then \mathbf{M} is not invertible because $\mathbf{M}\mathbf{e} = \mathbf{0}$.

Editorial comment. Many solvers noted that the same result holds if the condition $a_i \geq 0$ is weakened to $a_i \geq -1$.

Solved also by J. C. Binz (Switzerland), B. W. Brock, F. Brulois and T. Shore and D. B. Tyler, D. Callan, R. J. Chapman (U.K.), M. Dindos (Slovakia), Z. Franco, K. S. Kedlaya (student), N. Komanda, O. Krafft (Germany), O. P. Lossers (The Netherlands), M. Mócsy (Hungary), A. Nijenhuis, A. Pedersen (Denmark), E. Schmeichel, R. Stong, M. Tsatsomeris (Canada), and the proposer.

Collaborating editors: David F. Appleyard, Paul T. Bateman, Duane M. Broline, Barry W. Brunson, Frank S. Cater, Gulbank D. Chakerian, Underwood Dudley, Gerald A. Edgar, Michael A. Filaseta, Ira M. Gessel, Richard A. Gibbs, Jerrold R. Griggs, Douglas A. Hensley, John R. Isbell, Mourad E. H. Ismail, Murray Klamkin, Daniel J. Kleitman, Frederick W. Luttman, Frank B. Miles, Richard Pfiefer, Stephen L. Portnoy, J. O. Shallit, John Henry Steelman, Kenneth B. Stolarsky, David E. Tepper, Douglas B. Tyler, Daniel Ullman, and William E. Watkins.

Answer to Picture Puzzle
(p. 27)

Solomon Lefschetz.

REVIEWS

Edited by **Darrell Haile**
Indiana University, Bloomington, IN 47405

Mathematical Cranks by Underwood Dudley. The Mathematical Association of America, Washington, 1992, v + 372, \$25.00.

Reviewed by **Ian Stewart**

A few years ago I wrote a short article in *New Scientist*, in which I made passing reference to “the Fibonacci sequence 1, 1, 2, 3, 5, 8, 13, . . .” (The article was actually about the fractal structure of diffusion-limited aggregates, but that’s by the by.) Shortly afterwards the editors received a somewhat vituperative letter complaining about my appalling ignorance. I explained to the editors why I believed myself to be blameless, and they spiked the letter. Over the next few weeks they received repeated telephone calls from the same person, arguing the same point; he also tried to telephone me to discuss the matter, but my secretary managed to divert him with lies about my absence.

What was this terrible error that I had made, an error so important that the attempt to have it corrected occupied so much time and effort?

Simple. The Fibonacci series goes *zero*, 1, 1, 2, 3, 5, 8, 13, . . .

I don’t particularly want to argue the toss here, but I think most mathematicians would concur that whether you include an initial zero is largely a matter of convention. As it happened, I had excellent journalistic reasons for excluding it. The pattern of numbers then fitted the experimentalists’ picture precisely. If I had included the zero, I would also have had to spend several sentences explaining why it didn’t show up in the picture. All of this would be a distraction from the main point—about pattern-formation—and the word limit of 500 was already tight.

I don’t *know* that the person involved was a crank—but they exhibited one of the crucial symptoms, an obsession with trivialities. Very *specific* trivialities: I imagine he specialised in Fibonacci numerology. He was probably a perfectly reasonable person on most matters, but push the Fibonacci button and he would perform.

Cranks are a pest: to their families and friends, to us professionals, and to themselves. So why would anybody write an entire book about them? Because, says Underwood Dudley, ‘cranks are *interesting*,’ And he’s right.

Mathematical Cranks fits into a tiny but respected genre, which began with Augustus de Morgan’s *A Budget of Paradoxes*. Martin Gardner’s *Fads and Fallacies* is another, though not addressed to mathematics; and there’s also Dudley’s own *A Budget of Trisections*. Dudley collects cranks—it would be more prosaic, but less representative of the spirit of the enterprise, to say he collects crank *writings*—and he has solved a major problem for many of us. Before, when we received a crank letter, we would have to debate whether to answer it (and risk an endless and pointless ensuing correspondence), be exceedingly rude to its author in the hope that he (it virtually always is a *he*) will go away, or just bin it. (I must here confess to once binning a letter from a very nice doctor whose *patient* was the

crank. The doctor had no way of knowing whether the patient's mathematical research was valuable or not, and wanted professional advice. My problem was that the patient was at the time confined to a secure unit for the mentally ill that specialises in homicidal maniacs. Would *you* tell such a person—however nice and well-meaning his doctor is—that his life's work on Fermat's Last Theorem is complete trash?) Be that as it may, Dudley's hobby of crank-collection has solved all such dilemmas. Now I just shove the lot in an envelope and send it to *him*. I recommend that you all do the same. (He's at DePauw University, Greencastle, IN. No, look, he really *wants* this stuff. . . .)

The activities exposed in *Mathematical Cranks* take a variety of forms, ranging from utter nonsense to ideas that might—in other peoples' hands—form the basis of some genuine research projects. The Top Ten of mathematical crankhood is

- 1 Squaring the Circle
- 2 Trisecting the Angle
- 3 Fermat's Last Theorem
- 4 Proving the Parallel Postulate
- 5 The Golden Number
- 6 Perfect Numbers
- 7 The Four Colour Theorem
- 8 Number Bases
- 9 Cantor's Diagonal Argument
- 10 Duplicating the Cube.

It may seem strange that the 'three classical problems of antiquity' (actually nothing of the sort) come as numbers 1, 2, and then a huge gap until 10. But cranks' intuition agrees with that of most mathematicians: duplicating the cube is *boring* in comparison to the other two. Number base cranks advocate a change from base ten to something more sensible, usually 12 but sometimes 8 or 16. They are cranks not because this is a bad idea—on the contrary, base 12 would make excellent sense—but because as a matter of practicality it is even less likely to come about than a change from the QWERTY keyboard. (My own view here is that it would be simpler to reorder the alphabet to start at *Q*. Let me tell you all about it, it shouldn't take more than a week. . . .) The four-colour theorem has dropped down the charts since Appel-Haken: nowadays cranks confine themselves to seeking simple proofs—or at least proofs simpler than many hours of supercomputer time. Cantor's proof of the uncountability of the real numbers is perhaps the most sophisticated member of the Top Ten; Fermat's last theorem is the most significant open problem (added in proof: well, it was); and the parallel postulate has had the greatest influence on mathematics.

There are three patterns to this list. The first is that cranks have an unbounded belief in their own ability to solve problems for which there is a theorem that no solution exists. The second is that if a result is counter-intuitive, a crank will prefer to deny it rather than to refine his intuition. The third is that the kind of problem that cranks know about is the sort of thing that a schoolteacher might throw out to the odd bright pupil. While the rest of the class is struggling with how to bisect angles, there's one little lad with shining eyes who's set his sights on dividing them into n equal parts. Kindly teacher points out that the case $n = 3$ is known to be impossible; or perhaps just vaguely recalls being told that nobody has ever been able to solve it. At that instant a crank is born: a message is imprinted somewhere in the depths the child's brain. That message is *not* that the problem has been understood long ago: it is that it remains unsolved and that vast fame and fortune

await the intrepid solver. It is in the nature of the disease that it will then lie dormant for anything up to 50 years, to reveal its deadly symptoms only upon retirement, when the sufferer has finally found the time to work on the question that has been festering away these past many decades.

Teachers (and I here count myself as one): *do not worry about this*. We do far more good to the majority of our students by making it clear that mathematics includes both unsolved problems and unsolvable ones. We also do far more damage to our students in other unavoidable ways, and we usually never know it. That is in the nature of our chosen calling, just as doctors kill most of their patients in the long run. In any case, it seems likely that crankhood is in fact a genuine medical condition. There is a known disease of the mind—regrettably I can’t recall its name—whose symptoms include an obsession with very specific trivia. Cranks surely suffer from some version of this.

Some samples.

There is the fascinating case of HJ (Dudley uses initials only on the grounds that it is the ideas of cranks, not the people themselves, that he wishes to discuss) who invented Celestial Calculus. Ordinary calculus adds a little bit (δx) to x and sees by how much (δy) the function $y = f(x)$ changes. HJ *multiplies* x by something a little bit bigger than 1 and see how much y multiplies by. This is a neat idea. It replaces the usual differential operator $D = d/dx$ by a different one L , which can be expressed as

$$L(y) = \frac{d(\ln y)}{d(\ln x)}.$$

It’s rather cute:

$$L(x^n) = n$$

$$L(e^x) = x$$

$$L(uv^{\pm 1}) = L(u) \pm L(v).$$

The formula for $L(u + v)$ is not so cute, however, and $L(L(u))$ is a pig. A keen calculus class would have fun with this operator. What’s so cranky here? Nothing—until you read HJ’s estimation of its importance: “Clearly the most unifying and comprehensive identity in mathematics.”

WD simply couldn’t bear the Banach-Tarski Theorem, that a sphere can be cut into a finite number of pieces and reassembled to give two spheres the same size as the original. This would be paradoxical if it referred to physical spheres and to pieces that could actually be *made*, but it doesn’t and it’s not. WD realised that the proof of the Banach-Tarski Theorem depended on set theory, so he decided that set theory had to go. His chosen weak link was Cantor’s diagonal proof that the real numbers are uncountable, and his idea was to construct a countable list of all real numbers. Buried in the verbiage is the list: .1, .2, .3, .4, .5, .6, .7, .8, .9, .10, .11, .12, .13, . . . and so on. Students often make the same error, but they usually understand when it is pointed out that the list includes only the terminating decimals. Cranks don’t.

JB mailed a two-page paper to every math department in the United States, with the original and succinct title:

His discovery is the impressive-looking formula

$$\sum_{N^E}^{(N-1)^E} \left(E\sqrt{N^E} - \sqrt{N^E - 1} \right) = 1.$$

The notation's slightly adrift, and it's a useful exercise to correct it. He meant that, for example, when $N = 3$, $E = 2$ the sum should be

$$(\sqrt{9} - \sqrt{8}) + (\sqrt{8} - \sqrt{7}) + (\sqrt{7} - \sqrt{6}) + (\sqrt{6} - \sqrt{5}) + (\sqrt{5} - \sqrt{4}).$$

Think about it.

GB, some kind of engineer in his mid-fifties, spent a lot of time trying to find good ways to calculate tables of primes and the like. As a mathematician who knew him explains: "He genuinely expected some sort of ovation from the mathematical community, which he envisioned squirming in its chairs for want of good methods of calculating square roots, tabling Pythagorean triples, and listing the primes... What he has been doing lately is this: he considers the one-sided infinite matrix whose (i, j) entry is $i^2 + j$, and looks at the numbers that lie on lines, parabolas, and hyperbolas drawn in this array, with the idea of discovering patterns. He has discovered some God-awful patterns so far." For instance, the diagonal running upwards to the right from entry $(10, 1)$ reads 101, 83, 67, 53, 41, 31, 23, 17, 13, 11. All primes! One can see why GB was so interested. Unfortunately, because no polynomial can take only prime values, his search is ultimately doomed to failure.

Cranks have infinite confidence and zero modesty. LJ was so incensed about the primitive state of the theory of differential equations that he wrote *The Stupid, Ridiculous Oversight* to put the matter right. It is quite the exception among crank writing: it tackles such things as Yukawa's solution of the wave equation for the muon, and Schrödinger's equation for the energy levels of hydrogen. It contains an *exact* solution of van der Pol's equation, which LJ recasts as

$$y'' + Ay' + By^2y' + Cy = -Di e^{i\omega t}.$$

His solution takes the form

$$[\cos \omega t] \sum_{n=0}^{\infty} (\underline{n}/n!)(\omega t)^n.$$

We need to be told what \underline{n} is; and LJ explains in a series of equations that went

$$\begin{aligned} \underline{0} &= 1 \\ \underline{1} &= -(A + B)\omega^{-1}/2 - D\omega^{-2}/2 \end{aligned}$$

and so on until

$$\begin{aligned} \underline{4} &= \left[A(3 \underline{1} - \underline{3})\omega - B(\underline{3} - 27 \underline{1} + 6 \underline{12} + 2 \underline{111})\omega \right. \\ &\quad \left. - C(\underline{2} - 1)\right] \omega^{-2} + 6 \underline{2} - 1, \dots \end{aligned}$$

Dudley remarks that "the '...' doesn't help much, but never mind," and goes on to test the solution—whatever '...' means—by taking the extreme case $A = B = C = D = 0$. Now the equation becomes $y'' = 0$. The solution is

$$y = at + b$$

for constants a and b . LJ's solution, on the other hand, reduces to

$$y = (\cos \omega t) \left(1 + \frac{1}{2!} (\omega t)^2 + \frac{5}{4!} (\omega t)^4 + \dots \right).$$

"LJ must have committed an oversight, perhaps stupid and ridiculous."

As well as providing a representative survey of crank writings, and a generally compassionate but occasionally exasperated view of their authors, *Mathematical Cranks* also addresses various practical issues: how cranks are created, how to respond to them, and their (totally ungrounded) belief that it is possible to make money out of mathematics. In particular, we are urged *never* to encourage a crank with vague statements that *he* will read as praise but *you* know are the opposite—for example "If your work were to contain significant results, I would be happy to publish it in my journal." The crank doesn't see the 'if'—or if he does, he *knows* his results are significant. So why won't you publish them? Equally, it is highly dangerous to say "I found your results very interesting but I am not competent to judge them" when what you mean is "Your work is obvious trash but I'm not prepared to waste my time saying why". You may well find yourself quoted in the *next* version of the work—as being incompetent, but dazzled by its brilliance.

Mathematical Cranks is unique, and it exists. For both of these properties we should be grateful.

Mathematics Institute
University of Warwick
Coventry CV4 7AL, England

Complex Analysis: The Geometric Viewpoint. By Steven G. Krantz. The Carus Mathematical Monographs, Number 23, The Mathematical Association of America, ix + 212, 1990.

Reviewed by John Polking

One of the more fruitful and beautiful approaches to complex function theory in one and in several variables involves the interaction between holomorphic functions and mappings and various natural Riemannian metrics. The applications of this geometric approach include value distribution theory. The simplest example of this theory is the Liouville Theorem which says that any bounded entire function must be a constant. A more difficult example is the Little Picard Theorem, which says that any entire holomorphic function whose range omits two points must be a constant. Another application of the geometric approach is the study of properties of the group of biholomorphisms of a domain. These can range from determination of the group in simple cases such as the disk or the plane, to finding conditions under which the group is compact or noncompact.

Historically the first element of the geometric approach was provided by the introduction of what is called the Poincaré metric on the unit disk by Poincaré in 1881. This is the Riemannian metric

$$ds^2 = (1 - |z|^2)^{-2} (dx^2 + dy^2).$$

Using this metric the length of a curve parametrized by $\gamma(t)$ of $a \leq t \leq b$ is given by

$$L(\gamma) = \int_a^b \frac{|\gamma'(t)|}{1 - |\gamma(t)|^2} dt.$$

The Poincaré metric provides the disk with a geometric structure which is a model for noneuclidean geometry, a fact that is still a source of inspiration to geometers.

The application of complex analysis begins with Georg Pick's version of the standard Schwarz Lemma of complex analysis in 1916. This result, reinterpreted in modern geometric parlance, says that a holomorphic function f from the unit disk to itself cannot increase distance in the Poincaré metric. In terms of the length of curves it says that $L(f \circ \gamma) \leq L(\gamma)$. From this point on the geometric approach to complex function theory became a fruitful area of research.

A signal result in the development of the geometric approach was provided by Lars Ahlfors in 1938. He noticed that the Schwarz-Pick result remained true for maps from the disk equipped with the Poincaré metric to a domain U with an arbitrary metric of the form

$$ds^2 = \sigma(z)^2(dx^2 + dy^2), \quad (1)$$

as long as this metric has negative curvature, bounded above by -4 (the curvature of the Poincaré metric is identically equal to -4). "Curvature" refers to the ordinary curvature of Riemannian metric. In the case of metric like (1) a computation shows that it is equal to

$$\kappa_\sigma(z) = \frac{-\Delta \log \sigma(z)}{\sigma(z)^2}, \quad (2)$$

where

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

Over the past fifty years the geometric approach has seen remarkable advances in both one and several complex variables. It has proven to be a principal unifying theme in complex analysis.

As beautiful as this subject is, and as important as it has become in current research, there has not been an expository account available until the book by Steven Krantz appeared. Krantz's book is aimed at readers who have had one semester of complex analysis. Although geometry is used throughout, no prior knowledge of differential geometry is required to read it. Very little geometry is actually used, and that can be explained quite easily. The entire book can be considered to be an elementary approach to certain topics in complex function theory. In 200 small pages the geometric theory as outlined above (and more) is presented in a most coherent and clear manner. The book is a delight to read. It provides a wonderful advertisement for modern ideas in complex analysis and geometry.

The book begins (in Chapter 0) with a survey of the beginning of complex analysis. It is not a complete survey of that subject, since the author is interested in only part of it. Particular attention is paid to results involving restrictions on the range of a function such as the Liouville Theorem, the Casorati-Weierstrass

Theorem, and the Picard Theorems, and on results involving conformal mapping such as the Riemann Mapping Theorem and the classification of the biholomorphisms of the disk. The proofs here are not always complete, but they always convey the key ideas. In fact, foregoing complete proofs at times enables the author to provide a very nice overview of this part of complex analysis.

The basic geometric ideas that are needed in the book are introduced in Chapter 1. The only Riemannian metrics considered are those which are conformal to the euclidean metric, i.e. those of the form indicated in (1). Thus only the single coefficient function σ is of concern, greatly simplifying the geometrical considerations. This chapter ends with the Schwarz-Pick Lemma referred to earlier.

Chapter 2 begins with the introduction of the notion of curvature. The definition given is that in (2). Although this analytic formula is all that is needed, the formula seems unmotivated. The author realizes this, and includes a paragraph discussing the point. However this paragraph only directs the reader to the Appendix where there is a very formal discussion of curvature. Perhaps it would have been better to include a more informal discussion of curvature which would serve to convince the reader of the geometric content of the formula. The model of Chapter 0 could have been very appropriately followed here.

This minor quibble aside, Chapter 2 is the heart of the book. After the discussion of curvature, Ahlfors' version of the Schwarz Lemma is presented, followed fairly quickly by proofs of the Little Picard Theorem, a unifying discussion of the notion of the normal families using the spherical metric, and proofs of Montel's Theorem and the Great Picard Theorem. The quickness is due to the unifying effect of the Ahlfors result. For all of these applications one global result is needed—the rather easy fact that on the complex plane with two points removed, a metric can be found which has negative curvature bounded above by a constant $-B < 0$.

In Chapter 3 the author introduces the Carathéodory and Kobayashi/Royden metrics (in one variable). These are metrics which are defined on arbitrary domains in the complex plane (although they may degenerate in particular cases), and which are invariant under biholomorphisms. The principal applications are to the study of biholomorphism groups, and to normality of families of functions.

Finally in Chapter 4 the author turns to several complex variables. After some preliminaries, he introduces the Carathéodory and Kobayashi metrics in this context and uses the Kobayashi metric to prove that the unit ball and the bidisk in \mathbb{C}^2 are not biholomorphic. Since these domains are homeomorphic, this is in sharp contrast to the Riemann Mapping Theorem in one variable which says that any planar domain which is homeomorphic to the disk is biholomorphic to it. This is one example (out of many that are available) that shows that new and interesting phenomena occur in function theory in more than one complex variable.

To sum up, this is a fine book—one that continues the tradition of the Carus Mathematical Monographs for excellent exposition of important mathematics.

A final point about elementary textbooks in complex analysis. As the preceding discussion indicates, the geometric approach to function theory is too advanced to be given much coverage in a beginning, one semester course on complex analysis. The most one can expect at that level is a proof of the Schwarz Lemma, and the application to the classification of the biholomorphisms of the unit disk. However, given the importance of the geometric approach, at least this much should be included. An incomplete survey of beginning books on complex analysis reveals

that even that little is missing from many modern books. Some do not even mention the Schwarz Lemma. Others include it as a simple application of the maximum principle, but without any mention of applications. This kind of thing makes mathematics look like an exercise in mental gymnastics. Only about a third of the books examined contained both the Schwarz Lemma and its application. This lack of attention to a very important result is regrettable. A clear exception is the book of Ahlfors, which contains a discussion of the Lemma, followed by a proof, and then a series of eight wonderful exercises containing the application to the classification of biholomorphisms in addition to the connection with the Poincaré metric that is the result of the Schwarz-Pick Lemma.

Department of Mathematics
Rice University
P.O. Box 1892
Houston, TX 77251-1892



Submitted by Ralph Bean, Stockton State College, Pomona, New Jersey.

TELEGRAPHIC REVIEWS

Edited by Arnold Ostebee and Paul Zorn

with the assistance of the Mathematics Departments of
Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

<i>T</i> : Textbook	<i>P</i> : Professional Reading	1-4: Semester
<i>C</i> : Computer Software	<i>L</i> : Undergraduate Library	** : Special Emphasis
<i>S</i> : Supplementary Reading	13: Grade Level	?? : Questionable

Readers are advised that price information is subject to change. Selected books and software packages receive a second, more extensive review in the *Monthly*.

Books and software submitted for review should be sent to *Book Reviews Editor*, *American Mathematical Monthly*, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098.

Mathematics Appreciation, T(14-15: 1), S, C, L. *Exploring Mathematics With Your Computer*. Arthur Engel. New Math. Lib., V. 35. MAA, 1993, ix + 301 pp, \$38 (P). [ISBN 0-88385-639-5] A nice idea, not very well executed. Leaps into topics without much motivation. Still, there are mathematical nuggets worth panning. BC

History, S, L. *The Life of Isaac Newton*. Richard S. Westfall. Cambridge Univ Pr, 1993, xxii + 328 pp, \$24.95. [ISBN 0-521-43252-9] Condensed version of author's award-winning biography *Never at Rest: A Biography of Isaac Newton* (1980). This shorter, less technical version aims at a general audience. HD

Number Theory, P. *Algebraic Number Theory*. A. Fröhlich, M.J. Taylor. Stud. in Adv. Math., V. 27. Cambridge Univ Pr, 1993, xiv + 355 pp, \$29.95 (P). [ISBN 0-521-43834-9] An introduction to classical algebraic number theory, with some interesting twists. Covers standard topics, also cubic and biquadratic fields, elliptic curves, binary quadratic forms, Brauer relations. Some exercises. SG

Number Theory, P. *From Number Theory to Physics*. Eds: M. Waldschmidt, et al. Springer-Verlag, 1992, xiii + 690 pp, \$89. [ISBN 0-387-53342-7] Fourteen self-contained, expository lectures on many aspects of number theory. Most require no specialized background, but extend to recent developments in the field. A useful collection for any mathematician who wants to learn more about these fields. MPR

Number Theory, P. *A Tribute to Emil Grosswald: Number Theory and Related Analysis*.

Eds: Marvin Knopp, Mark Sheingorn. *Contemp. Math.*, V. 143. AMS, 1993, viii + 612 pp, \$79 (P). [ISBN 0-8218-5155-1] 39 articles on topics in number theory, modular functions, combinatorics, and analysis.

Linear Algebra, T(14: 1). *Introduction to Linear Algebra*. Gilbert Strang. Wellesley-Cambridge Pr, 1993, viii + 472 pp, \$42. [ISBN 0-9614088-5-5] Basic philosophy same as author's *Linear Algebra and Its Applications* (TR, August-September 1988), but moves more slowly. Covers standard topics plus applications, numerical linear algebra. Includes some MATLAB programs. LC

Algebra, P. *General Algebra and Applications*. Eds: K. Denecke, H.-J. Vogel. Res. & Expos. in Math., V. 20. Heldermann Verlag, 1993, 237 pp, DM 78 (P). [ISBN 3-88538-220-2] 20 articles based on lectures given at the "43. Arbeitstagung Allgemeine Algebra" (Potsdam, 1992).

Algebra, P. *Representation Theory of Groups and Algebras*. Eds: Jeffrey Adams, et al. *Contemp. Math.*, V. 145. AMS, 1993, x + 491 pp, \$50 (P). [ISBN 0-8218-5168-3] Proceedings of the Representation Theory Group at the University of Maryland during the years 1989-1992. Lecture notes from keynote speakers as well as research and expository articles.

Complex Analysis, P. *Several Complex Variables in China*. Eds: Chung-Chun Yang, Sheng Gong. *Contemp. Math.*, V. 142. AMS, 1993, xii + 173 pp, \$36 (P). [ISBN 0-8218-5164-0] 9 papers, several in survey style, on topics in several complex variables (singular integrals,

function spaces, differential operators, factorization, biholomorphic mappings, etc.). PZ

Complex Analysis, P. *Several Complex Variables: Proceedings of the Mittag-Leffler Institute, 1987–1988.* Ed: John Erik Fornæss. Math. News, V. 38. Princeton Univ Pr, 1993, vii + 701 pp, \$39.50 (P). [ISBN 0-691-08579-X] 38 papers by participants in the special year devoted to several complex variables.

Differential Equations, S(17–18), P. *Oscillations in Nonlinear Systems.* Jack K. Hale. Dover, 1992, ix + 180 pp, \$7.95 (P). [ISBN 0-486-67362-6] Periodic and almost-periodic solutions to non-linear ODE's with a small parameter. Treats integral manifolds, method of averaging. SK

Partial Differential Equations, T(17–18: 2–4), L. *An Introduction to Partial Differential Equations.* Michael Renardy, Robert C. Rogers. Texts in Appl. Math., V. 13. Springer-Verlag, 1993, xiii + 428 pp, \$42. [ISBN 0-387-97952-2] Text for a broad-based introductory graduate course in PDE's. JO

Functional Analysis, P. *Banach Spaces.* Eds: Bor-Luh Lin, William B. Johnson. Contemp. Math., V. 144. AMS, 1993, xviii + 201 pp, \$42 (P). [ISBN 0-8218-5157-8] Proceedings of January 1992 workshop at the Universidad de Los Andes in Merida, Venezuela.

Analysis, T(17–18: 1), P. *Spline Functions and Multivariate Interpolations.* B.D. Bojanov, H.A. Hakopian, A.A. Sahakian. Math. & Its Applic., V. 248. Kluwer Academic, 1993, ix + 276 pp, \$122.50. [ISBN 0-7923-2229-0] Comprehensive introduction to the theory of spline functions, stressing new developments. Some results appear as exercises, with generous hints. Historical notes, extensive bibliography, relatively few exercises. HD

Analysis, P. *Representation of Lie Groups and Special Functions, Volume 2: Class I Representations, Special Functions, and Integral Transforms.* N. Ja. Vilenkin, A.U. Klimyk. Transl: V.A. Groza, A.A. Groza. Math. & Its Applic., V. 74. Kluwer Academic, 1993, xviii + 607 pp, \$290. [ISBN 0-7923-1492-1] Second of three volumes treats (i) properties of special functions and orthogonal polynomials that are related to the class I representations of certain groups; (ii) q -analogs of classical orthogonal polynomials related to representations of the Chevalley groups; (iii) special functions connected to fields of p -adic numbers. LC

Differential Geometry, T(15–16: 1, 2), S, L*. *Curves and Singularities: A Geometrical Introduction to Singularity Theory, Second Edition.* J.W. Bruce, P.J. Giblin. Cambridge Univ

Pr, 1992, xviii + 321 pp, \$69.95. [ISBN 0-521-41985-9] This delightfully written book overflows with beautiful mathematics, requiring only linear algebra, multi-variable calculus, and a little mathematical sophistication. Buy a copy for a friend. JO

Differential Geometry, P. *The Problem of Plateau—A Tribute to Jesse Douglas & Tibor Radó.* Ed: Themistocles M. Rassias. World Scientific, 1992, x + 335 pp, \$94. [ISBN 981-02-0556-2] First section recounts lives and works of Plateau, Douglas, and Radó. Second collects recent work on Plateau's Problem. JO

Geometry, T(18: 2). *Convex Bodies: The Brunn-Minkowski Theory.* Rolf Schneider. Ency. of Math. & Its Applic., V. 44. Cambridge Univ Pr, 1993, xiii + 490 pp, \$89.95. [ISBN 0-521-35220-7] Basic properties and boundary structure of convex bodies, Minkowski addition, curvature measures, basic properties of mixed volume and mixed area measures, inequalities satisfied by mixed volumes. RWJ

Control Theory, P. *Lecture Notes in Control and Information Sciences-185: Analysis and Optimization of Systems: State and Frequency Domain Approaches for Infinite-Dimensional Systems.* Eds: R.F. Curtain, A. Bensoussan, J.L. Lions. Springer-Verlag, 1993, xiii + 648 pp, (P). [ISBN 0-387-56155-2] Proceedings of the 10th International Conference on Analysis and Optimization of Systems (Sophia-Antipolis, France, 1992).

Control Theory, P. *Lecture Notes in Control and Information Sciences-161: Topics in Stochastic Systems: Modelling, Estimation and Adaptive Control.* Eds: L. Gerencsér, P.E. Caines. Springer-Verlag, 1991, 401 pp, \$61 (P). [ISBN 0-387-54133-0] 16 survey papers on recent work and trends.

Probability, P. *Spatial Stochastic Processes: A Festschrift in Honor of Ted Harris on his Seventieth Birthday.* Eds: Kenneth S. Alexander, Joseph C. Watkins. Progress in Prob., V. 19. Birkhäuser, 1991, xii + 256 pp, \$84. [ISBN 0-8176-3477-0] 11 papers on branching processes, percolation, interacting particle systems, and stochastic flows.

Stochastic Processes, P. *Selected Works of A.N. Kolmogorov, Volume III: Information Theory and the Theory of Algorithms.* Ed: A.N. Shiryaev. Transl: A.B. Sossinsky. Math. & Its Applic., V. 27. Kluwer Academic, 1993, xxv + 275 pp, \$132. [ISBN 90-277-2798-8] A lengthy biography, then twelve major papers on information theory, ϵ -entropy and capacity, and algorithm analysis. TAV

Statistics, T(16–17: 2), L. *Introduction to*

Probability and Statistics, Second Edition, Revised and Expanded. Narayan C. Giri. Stat.: Textbooks & Mono., V. 136. Marcel Dekker, 1993, xv + 537 pp, \$59.75. [ISBN 0-8247-9037-5] Combines *First Edition* volumes on probability (*Part I*, TR, November 1975) and statistics (*Part II*, TR, March 1976). New edition expands coverage of multivariate distributions, the normal in particular. Added topics include decision theory and the application of ANOVA to experimental design. RWJ

Statistics, T(16-17: 1). *Statistical Methods in Agriculture and Experimental Biology, Second Edition.* R. Mead, R.N. Curnow, A.M. Hasted. Chapman & Hall, 1993, xi + 415 pp, \$44.95 (P); \$99.95. [ISBN 0-412-35480-2; 0-412-35470-5] New edition includes topics uncommon in introductory texts: modeling, analysis of multiple measurements, etc. Mathematically elementary. (*First Edition*, TR, April 1984.) KB

Statistics, P. *Pitman's Measure of Closeness: A Comparison of Statistical Estimators.* Jerome P. Keating, Robert L. Mason, Pranab K. Sen. SIAM, 1993, xx + 226 pp, \$26.50 (P). [ISBN 0-89871-308-0] Nicely presents history of Pitman's measure of closeness (PMC), applications to single-parameter estimation problems, PMC anomalies, and asymptotics. RWJ

Statistics, T(14-15: 1). *Applied Nonparametric Statistical Methods, Second Edition.* P. Sprent. Chapman & Hall, 1993, x + 342 pp, \$34.50 (P). [ISBN 0-412-44980-3] More emphasis on computing and links between apparently disparate procedures. New material on categorical data analysis. (*First Edition*, TR, November 1989). KB

Algorithms, T(16), S, L. *The Design and Analysis of Parallel Algorithms.* Justin R. Smith. Oxford Univ Pr, 1993, xviii + 510 pp, \$75. [ISBN 0-19-507881-0] Stresses SIMD algorithms, data level parallelism. Examples from numerical and scientific computation, symbolic algorithms. Discusses models (PRAM, distributed memory), design principles, existing systems, the language C*. RM

Algorithms, P. *Algorithms: Main Ideas and Applications.* Vladimir Uspensky, Alexei Semenov. Math. & Its Applic., V. 251. Kluwer Academic, 1993, xii + 269 pp, \$116. [ISBN 0-7923-2210-X] Fundamental, theoretical treatment of the general theory of algorithm, motivated by mathematical and philosophical foundations, with attention to historical development. RM

Computer Graphics, T(16-17: 1, 2), C, P, L. *Curves and Surfaces for Computer Aided*

Geometric Design: A Practical Guide, Third Edition. Gerald Farin. Comp. Sci. & Sci. Comp. Academic Pr, 1993, xvii + 473 pp, \$49.95. [ISBN 0-12-249052-2] Main ideas of Computer Aided Geometric Design; a chapter by Bezier treats origins of Bezier splines. Included disk contains C programs, PostScript files. (*Second Edition*, TR, January 1991.) JO

Computer Science, T(17: 1), P. *String-Rewriting Systems.* Ronald V. Book, Friedrich Otto. Texts & Mono. in Comp. Sci. Springer-Verlag, 1993, viii + 189 pp, \$39. [ISBN 0-387-97965-4] String-rewriting systems, introduced by Thue in 1914 for discussion of "word-type" problems, have been applied to theory of formal languages, more general rewriting problems, automated deduction, and algorithmic problems related to algebraic structures. Compiles recent results in the field. RM

Computer Science, P. *Logic Programming Languages.* Eds: K.R. Apt, J.W. de Bakker, J.J.M.M. Rutten. MIT Pr, 1993, xiv + 204 pp, \$32.50. [ISBN 0-262-01134-4] Research results from a project funded by ESPRIT, a European multinational organization, that aims to integrate logic and functional programming. RJA

Applications (Physical Science), P. *Air Pollution Modeling and Its Application IX.* Eds: Han van Dop, George Kallos. NATO Challenges of Mod. Soc., V. 17. Plenum Pr, 1992, xii + 803 pp, \$165. [ISBN 0-306-44248-5] Proceedings of the 19th NATO/CCMS International Technical Meeting on Air Pollution Modeling and Its Application (Crete, Greece, 1991).

Applications, P. *Biological Neural Networks in Invertebrate Neuroethology and Robotics.* Eds: Randall D. Beer, Roy E. Ritzmann, Thomas McKenna. Neural Networks: Found. to Applic. Academic Pr, 1993, xi + 417 pp, \$64.95. [ISBN 0-12-084728-0] 17 papers from a 1991 ONR workshop for invertebrate neuroethologists studying sensorimotor control and engineers studying autonomous mobile robots. Aims to apply biological knowledge and examples to engineering problems. Papers treat control of leg movement, orientation, computer modeling, robotics. RM

Reviewers

RJA: Richard J. Allen, St. Olaf; KB: Karla Ballman, Macalester; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; HD: Hung Dinh, Macalester; SG: Steven Galovich, Carleton; RWJ: Roger W. Johnson, Carleton; SK: Steve Kennedy, St. Olaf; RM: Richard Molnar, Macalester; JO: Jeff Ondich, Carleton; MPR: Matthew P. Richey, St. Olaf; TAV: Theodore A. Vessey, St. Olaf; PZ: Paul Zorn, St. Olaf.

Symbolic Computation in Undergraduate Mathematics Education

Zaven Karian, Editor

If you are considering putting a symbolic computing system into your curriculum, this is one publication you should have.

—Mathematics Teacher

This well-written book should be helpful to anyone using symbolic computation as an aid in teaching undergraduates—The book provides a number of examples for presenting probability and statistics in a way that removes the tedium and emphasizes the underlying ideas.

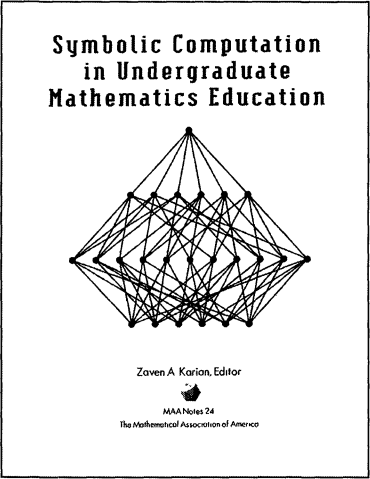
—AAAS, Science Book and Films

If you have any plans to integrate symbolic computing into your program, read and study this book first. Your students will thank you for it.

—AMATYC Review

This volume brings together many of the facets associated with the pedagogic uses of symbolic computation.

Part I consists of articles that deal with general issues of learning mathematics and the role of symbolic computation in that process. The articles in Part II describe the use of symbolic computation in teaching calculus. Some of the areas covered are the use of symbolic computation in a laboratory calculus course, the uses of Derive in the instruction of calculus, antidifferentiation and the



definite integral, and the experiences and reflections of teachers who have used symbolic computation in calculus instruction.

Part III consists of papers on sophomore-level courses on linear algebra and differential equations. The articles in Part IV describe what can be done in using symbolic computation in teaching combinatorics, probability and statistics courses. The articles and references in Part V will help you get started in using some of these ideas at your own institution.

200 pp., 1992, Paperbound
ISBN 0-88385-082-6
List: \$22.00
Catalog Number NTE-24

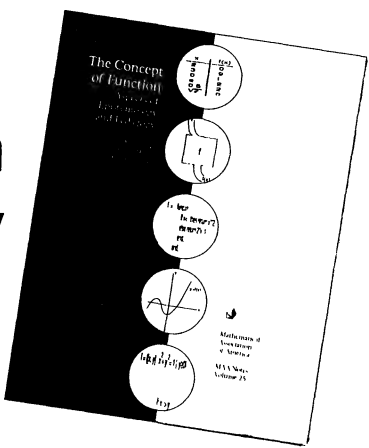
ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 2003
1-800-331-1622 Fax (202) 265-2384

Membership Code	Qty.	Catalog Number	Price
-----	_____	_____	_____
Name _____	_____		
Address _____	_____		
City _____	Total \$ _____		
State _____ Zip Code _____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
	Credit Card No. _____		
	Signature _____ Exp. Date _____		

The Concept of Function Aspects of Epistemology and Pedagogy

Guershon Harel and Ed Dubinsky, Editors



The contributors of this volume probe the idea of what it means to learn the concept of function and how instruction, based on research, could assist teachers in finding ways of helping their students understand this all-important mathematical concept.

The concept of function is one that will appear again and again in a student's mathematics training. Arithmetic in the early grades, algebra in junior high school, and transformational geometry in high school are all largely based on the idea of function. Moreover, people involved in calculus reform know that understanding the idea of function is an indispensable part of the background students need to understand calculus. As mathematical education is being renewed and reformed throughout the world, this movement requires that we learn more about the concept of function both from epistemological and pedagogical points of view.

There are several major themes that emerge in the pages of this volume. They are theoretical perspectives of development of the function concept, theory-based teaching experiments, conceptions held by students and teachers, and the use of pedagogical software. The volume begins

with a summary and overview of the subject and is followed by a brief glossary of terms.

The development of the papers presented in the volume began with a conference held in West Lafayette, Indiana in October 1990 with the support of Purdue University and the Exxon Foundation. This volume is, however, much more than just a conference proceedings. It is a truly cooperative writing effort by a group of dedicated researchers and educators.

350 pp., 1992, Paperbound
ISBN 0-88385-081-8

List: \$22.00

Catalog Number NTE-25

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
(202) 387-5200 Fax (202) 265-2384

	Qty.	Catalog Number	Price
Name _____			
Address _____			
City _____			
State _____ Zip Code _____			
			Total \$ _____
			Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
			Credit Card No. _____
			Signature _____ Exp. Date _____

JOURNEY INTO GEOMETRIES

Marta Sved

Foreword by H.S.M. Coxeter

All prospective teachers of secondary mathematics should be required to study this book—both content and style. —CHOICE

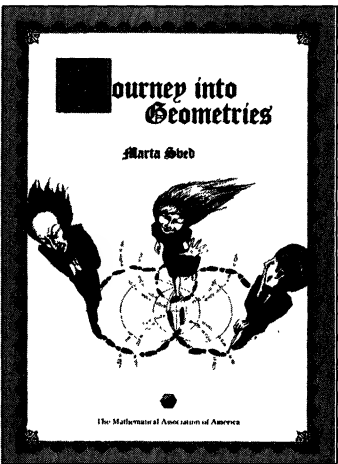
...an excursion into the wonderland of advanced geometry. Experts will frolic, amateurs will grin and eventually understand, and beginners will smile right up to the point (pun intended) where they get lost. A joy to read and even better the second time through, it is a fitting sequel to the original adventures.

—Journal of Recreational Mathematics

This is definitely a book for the school library, and for the teacher's shelf too.

—The Australian Mathematics Teacher

This charming book introduces us to topics in hyperbolic geometry in a delightfully informal style. Early in the 19th century, Janos Bolyai created "non-Euclidean" geometry, discovered independently by two other mathematicians of Bolyai's day, Gauss, and Lobachevsky. At the time these concepts were too revolutionary to make a serious impact. However, later developments in relativity theory and twentieth century perceptions made hyperbolic geometry an integral part of geometry, logically as perfect as classical geometry, yet still strangely surprising.



Journey into Geometries can be read at two levels. It can be studied as an informal introduction to post-Euclidean geometry, brought to life in dialogues between three fictitious figures: a somewhat grown up Alice, Lewis Carroll and their visitor from the Twentieth century, Dr. Whatif. It also can serve as background material for university students, for the material presented in the text is extended by carefully selected problems. The background required is minimal, standard high school geometry, yet the serious student, aided by problems attached to each chapter, should acquire a deeper understanding of the subject.

192 pp., Paperbound, 1991
ISBN 0-88385-500-3
List: \$22.00 MAA Member: \$16.00
Catalog Number JOG

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

Membership Code	Qty.	Catalog Number	Price

Name _____			
Address _____			
City _____			Total \$ _____
State _____ Zip Code _____			Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
			Credit Card No. _____
			Signature _____ Exp. Date _____

LURE OF THE INTEGERS

Joe Roberts

A joy to read and ponder, this book is a welcome addition to the body of mathematical literature. It belongs in every mathematical library.

—*Journal of Recreational Mathematics*

Will enrich library collections serving curricula with theory of numbers courses. —*Choice*

In some small way, this book is an introduction to a mythical book which might go under the name of *The Book of Integers*. This mythical book has on page n all of the interesting properties of the integer n . This introduction stems from many years' casual accumulation of numerical facts. Most of the material presented belongs to elementary mathematics in the sense that no deep or profound mathematical background is required in order to understand what is said. Much of the material is drawn from the theory of numbers.

Many of the topics touch on contemporary research and most of the results are stated without proof. As a general rule, one cannot tell from the statements of the results whether or not their proofs will be elementary. Indeed, this is a hallmark of mathematics and is one of the things that gives the subject a special flavor and interest. Until one knows that expert practitioners have been unable to solve a problem, one does not know that the problem is difficult. Even then it may turn out that there is an easy solution.

Some of the material will be familiar to people having only a small acquaintance with mathematics. Even in those cases, the author provides something new. On the other hand, much of the material is sufficiently out of the main stream of concern that even professional mathematicians may be unfamiliar with the results. The many references to the literature will almost always enable a reader to track down further information. In **Lure of the Integers** the author has presented a body of material which will prove interesting to the enlightened layman as well as to the professional.

300 pp., Paperbound, 1992

ISBN-0-88385-502-X

List: \$28.50 MAA Member: \$19.50

Catalog Number LURE

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

Foreign Orders Please add \$3.00 per item ordered to cover postage and handling fees. The order will be sent via surface mail. If you want your order sent by air, we will be happy to send you a proforma invoice for your order.

Membership Code _____	Qty.	Catalog Number	Price
Name _____	_____		
Address _____	_____		
City _____	Total \$ _____		
State _____ Zip Code _____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
	Credit Card No. _____		
	Signature _____ Exp. Date _____		

The Search for E.T. Bell

also known as John Taine

Constance Reid

No one today writes about mathematics and mathematicians with more grace, knowledge, skill, and clarity, and no one is going to produce a more delightful, informative, accurate account of Eric Temple Bell and his work, and that of his alter-ego, the prolific pioneer of science fiction, John Taine. This is a fine book. —Martin Gardner

Eric Temple Bell has been one of my heroes for 60 years...I congratulate Constance Reid on a remarkable achievement. I hope it is greeted with the success it deserves, and revives interest in an extraordinary and multi-talented man. —A. C. Clarke

Eric Temple Bell (1883–1960) was a distinguished mathematician and a best selling popularizer of mathematics. His *Men of Mathematics*, still in print after almost sixty years, inspired scores of young readers to become mathematicians. Under the name "John Taine," he also published science fiction novels (among them *The Time Stream*, *Before the Dawn*, and *The Crystal Horde*) that served to broaden the subject matter of that genre during its early years.

In *The Search for E.T. Bell*, Constance Reid has given us a compelling account of this complicated, difficult man who never divulged to anyone, not even to his wife and son, the story of his early life and family background. Her book is thus more of a mystery than a traditional biography. It begins with the discovery of an unexpected inscription in an English churchyard and a series of cryptic notations in a boy's schoolbook. Then comes an

inadvertent revelation, by Bell himself, in a respected mathematical journal...You will have to read the book to learn the rest.

Originally agreeing to write only a profile of Bell, Mrs. Reid soon found herself involved in a full-length biography. The discoveries she made in the course of her five years of research will necessitate a fresh evaluation of his extensive mathematical work and his science fiction novels as well as the revision of almost every statement currently in print about his family background and early life. Mrs. Reid is already well known as the author of acclaimed biographies of David Hilbert, Richard Courant, and Jerzy Neyman.

Includes a collection of over 75 photographs.

384 pp., Hardbound, 1993

ISBN 0-88385-508-9

List: \$35.00 MAA Member: \$28.00

Catalog Number BELL

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
(202) 387-5200 1-(800)331-1622

Membership Code

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty.	Catalog Number	Price
------	----------------	-------

_____	_____	_____
_____	_____	_____

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

Game Theory and Strategy

Philip D. Straffin, Jr.

This valuable addition to the New Mathematical Library series pays careful attention to applications of game theory in a wide variety of disciplines. The applications are treated in considerable depth. The book assumes only high school algebra, yet gently builds to mathematical thinking of some sophistication. **Game Theory and Strategy** might serve as an introduction to both axiomatic mathematical thinking and the fundamental process of mathematical modelling. It gives insight into both the nature of pure mathematics, and the way in which mathematics can be applied to real problems.

Since its creation by John von Neumann and Oskar Morgenstern in 1944, game theory has contributed new insights to business, politics, economics, social psychology, philosophy, and evolutionary biology. In this book, the fundamental ideas of game theory share the stage with applications of the theory. How might strategic business decisions depend on information about a rival company, and how much would such information be worth? When is it advantageous to vote for a candidate who is not your favorite? What are the optimal strategies for teams in the football draft, and what paradoxes can result from following

those strategies? What is a fair way to share the costs of a development project? What can we learn about the problem of "free will" by imagining playing a game with an omnipotent Being? How might natural selection lead to altruistic behavior in animal species? Game theory gives insight into all of these questions.

The book includes many exercises, with answers, which allow the reader to try out calculations, and explore alternative formulations of game-theoretic ideas.

200 pp., 1993, Paperbound

ISBN 0-88385-637-9

List: \$27.50 MAA Member: \$22.00

Catalog Number NML-36

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
(202) 387-5200 1-(800) 331-1622

Membership Code	Qty.	Catalog Number	Price

Name _____			
Address _____			Total \$ _____
City _____			Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
State ____ Zip Code _____			Credit Card No. _____
			Signature _____
			Exp. Date _____

Proofs Without Words

Exercises in Visual Thinking

Roger B. Nelsen

Just what are “proofs without words?” First of all, most mathematicians would agree that they certainly are not “proofs” in the formal sense. Indeed, the question does not have a simple answer. Proofs without words are generally pictures or diagrams that help the reader see *why* a particular mathematical statement may be true, and *how* one could begin to go about proving it. While in some proofs without words an equation or two may appear to help guide that process, the emphasis is clearly on providing *visual* clues to stimulate mathematical thought. Proofs without words bear witness to the observation that often in the English language to *see* means to *understand*, as in “to see the point of an argument.”

Proofs without words have a long history. In this collection you will find modern renditions of proofs from ancient China, classical Greece, twelfth-century India—even one based on a published proof by a former President of the United States! However, most of the proofs are more recent creations, and many are taken from the pages of MAA journals.

The proofs in this collection are arranged by topic into six chapters: Geometry and Algebra; Trigonometry,

Calculus and Analytic Geometry; Inequalities; Integer Sums; Sequences and Series; and Miscellaneous. Teachers will find that many of the proofs in this collection are well suited for classroom discussion and for helping students to think visually in mathematics.

The readers of this collection will find enjoyment in discovering or rediscovering some elegant visual demonstrations of certain mathematical ideas that teachers will want to share with their students. Readers may even be encouraged to create new “proofs without words.”

160 pp., Paperbound, 1993

ISBN 0-88385-700-6

List: \$27.50 MAA Member: \$22.00

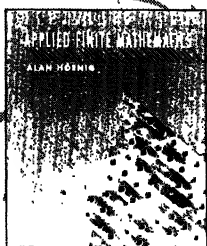
Catalog Number PWW

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
(202) 387-5200 1-800-331-1622

-----		Qty.	Catalog Number	Price
Membership Code -----		-----		
Name _____		Total \$ _____		
Address _____		Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
City _____		Credit Card No. _____		
State _____ Zip Code _____		Signature _____		
		Exp. Date _____		

Count on Houghton Mifflin for the finest in math textbooks



Applied Finite Mathematics, Second Edition **NEW!**

Alan Hoenig, John Jay College, City University of New York

Hoenig offers comprehensive, balanced coverage of the math topics that business, economics, and life- and social-science majors need to know. Clear, careful exposition of the topics, a wealth of real-world applications, and continuous skills practice make this text outstanding.

740 pages • hardcover • complete support package • just published (©1994)



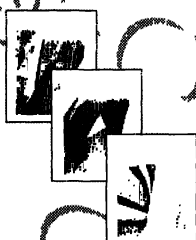
Calculus and Its Applications

Brief Calculus and Its Applications

Daniel D. Benice, Montgomery College

These texts provide a clear conceptual framework that reveals the "hows" and "whys" of calculus, with excellent exercises and realistic applications. Students develop their critical-thinking skills through writing exercises and an emphasis on problem-solving strategies.

Both hardcover • complete support packages • now available for review (©1993)



Fundamentals of College Mathematics **NEW!**

Introduction to Algebra

Intermediate Algebra

Sandra Clarkson and Barbara Barone

Both of Hunter College, City University of New York

All paperback • complete support packages • just published (©1994)



Prealgebra **NEW!**

Richard N. Aufmann, Palomar College

Vernon C. Barker, Palomar College

Joanne Lockwood, Plymouth State College

672 pages • paperback • complete support package • just published (©1994)

Algebra for College Students: A Functions Approach **NEW!**

Richard N. Aufmann, Palomar College

Joanne Lockwood, Plymouth State College

768 pages • hardcover • complete support package • just published (©1994)



Business Mathematics, Second Edition **NEW!**

Richard N. Aufmann, Palomar College

Vernon C. Barker, Palomar College

Joanne Lockwood, Plymouth State College

704 pages • paperback • complete support package • just published (©1994)

Also available in a Brief Edition

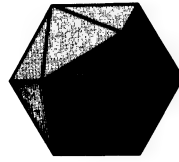
384 pages • paperback • complete support package • just published (©1994)

**Houghton
Mifflin**

To request an examination copy, contact your Houghton Mifflin sales representative or call 1-800-733-1717. In Canada, call 1-800-268-4404.

The American Mathematical Monthly

Volume 101 Number 1 / JANUARY 1994
(ISSN 0002-9890)



Contents

ARTICLES

Introduction to Fermat's Last Theorem / DAVID COX 3

Future Elementary Teachers: The Neglected Constituency /
THOMAS W. HUNGERFORD 15

Galois Theory for Beginners / JOHN STILLWELL 22

Into the Hourglass: Reflections on the Forces Acting on a Granular
Material / E. BRUCE PITMAN 28

Orderly Currencies / JOHN DEWEY JONES 36

An Interior Fixed Point Property of the Disc / ROBERT F. BROWN
and ROBERT E. GREENE 39

Le Cam's Inequality and Poisson Approximations /
J. MICHAEL STEELE 48

FEATURES

COMMENTS 2

PICTURE PUZZLE 27

NOTES

A Generalization of a Theorem of Euler / DORINA MITREA
AND MARIUS MITREA 55

The Existence of a Triangle with Prescribed Angle Bisector Lengths /
PETRU MIRONESCU and LAURENTIU PANAITOPOL 58

THE COMPUTER SCIENCE SAMPLER

Turing Machines and Computational Complexity / BILL MARION 61

THE EVOLUTION OF ...

The Evolution of Integration / A. SHENITZER and J. STEPRĀNS 66

THE AUTHORS 73

PROBLEMS AND SOLUTIONS 75

REVIEWS

Mathematical Cranks, by Underwood Dudley / IAN STEWART 87

Complex Analysis: The Geometric Viewpoint, by Steven G. Krantz /
JOHN POLKING 91

TELEGRAPHIC REVIEWS 95

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036



The American Mathematical Monthly



Volume 101, Number 2 / FEBRUARY 1994



AN OFFICIAL PUBLICATION OF THE MATHEMATICAL ASSOCIATION OF AMERICA

NOTICE TO AUTHORS

The *Monthly* publishes articles, notes, and other features about mathematics and the profession. The readership of the *Monthly* is intended to include everybody who is mathematically inclined, including of course professional mathematicians and students of mathematics at all collegiate levels. While no single article or feature is likely to appeal to everyone, material should interest and be accessible to a large number of readers. This is the most important criterion for acceptance.

Articles may be expositions of old results or presentations of new ones. They may concern all of mathematics or one small area, a broad development or a single application, historical reminiscences or one important event. While some articles may contain the author's new research, the novelty of material and generality of the results is far less important than the clarity of exposition and general interest. Discussing one illuminating case of a well known result is far better than providing all the details of an obscure but new proposition. Articles in the *Monthly* are supposed to inform and to entertain; they are meant to be read rather than archived.

Notes are short and possibly informal articles. A note may concern a clever new proof of an old theorem, a novel way to present tired material, or a lively discussion of a philosophical (but still mathematical) issue. Also, any topic is suitable, so long as it is related to mathematics. Because a note is short, the first few sentences are the most important part: They should explain the purpose and invite the reader in. Photographs or diagrams often will attract the reader's attention.

All articles and notes should be sent to the editor:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Please send 3 copies, typewritten on only one side of the paper. Illustrations should be carefully drawn on separate sheets of paper in black ink; the original should be without lettering and two copies should have appropriate captions and lettering indicated.

Proposed problems or solutions should be sent to:

RICHARD BUMBY,
P.O. Box 10971
New Brunswick, NJ 08906-0971.

Please send 2 copies of all material, typewritten if possible.

Letters to the Editor, both for publication and for private reading, should be sent to the Editor at the address given above. Comments, including criticisms, are welcome, as are all suggestions for making the *Monthly* a lively, entertaining, and informative journal.

EDITOR:

JOHN H. EWING

ASSOCIATE EDITORS:

RONALD BOOK	JOAN HUTCHINSON
PETER BORWEIN	CATHERINE MCGEOCH
RICHARD BUMBY	RICHARD NOWAKOWSKI
DENNIS DETURCK	ARNOLD OSTEBEE
UNDERWOOD DUDLEY	LEE RUBEL
JOHN DUNCAN	LYNN STEEN
JOAN FERRINI-MUNDY	STAN WAGON
JOSEPH GALLIAN	DOUGLAS WEST
STEVEN GALOVICH	HERBERT WILF
RICHARD GUY	SANDY ZABELL
DARRELL HAILE	PAUL ZORN
PAUL HALMOS	

EDITORIAL ASSISTANT:

MISTY CUMMINGS

STAFF ARTIST:

MIKE CAGLE

Reprint permission:

MARCIA P. SWARD, Executive Director

Advertising Correspondence:

Ms. ELAINE PEDREIRA, Advertising Manager

Subscription correspondence, change of address, and other inquiries:

Membership / Subscriptions Department

All at the address:

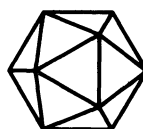
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036.

Microfilm Editions: University Microfilms International,
Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT. Copyrighted by the Mathematical Association of America (Incorporated), 1994, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source. Second class postage paid at Washington, DC, and additional mailing offices. **Postmaster:** Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.

**The American
Mathematical Monthly**

Volume 101 Number 2 / FEBRUARY 1994
(ISSN 0002-9890)



Contents

ARTICLES

Yueh-Gin Gung and Dr. Charles Y. Hu Award for Distinguished Service to
J. Sutherland Frame / DAVID W. BALLEW 107

A New Look at Euler's Theorem for Polyhedra / BRANKO GRÜNBAUM
and G. C. SHEPHARD 109

Otto Neugebauer: Reminiscences and Appreciation / PHILIP J. DAVIS
129

From the Buffon Needle Problem to the Kreiss Matrix Theorem /
ELIAS WEGERT and LLOYD N. TREFETHEN 132

A Counterexample for Germain / WILLIAM C. WATERHOUSE 140

Cubic Equations, or Where Did the Examination Question Come From?/
H. B. GRIFFITHS and A. E. HIRST 151

FEATURES

COMMENTS 106

PICTURE PUZZLE 139

NOTES

On the Identity of Polyhedra / HELLMUTH STACHEL 162

More on the Pompeiu Problem / DAVID C. ULLRICH 165

UNSOLVED PROBLEMS

Every Number Is Expressible as the Sum of How Many
Polygonal Numbers? / RICHARD K. GUY 169

THE AUTHORS 173

PROBLEMS AND SOLUTIONS 175

REVIEWS

Reality Rules I. The Fundamentals; II. The Frontier. By John Casti /
RUTHERFORD ARIS 186

Geometry of Surfaces. By John Stillwell / DAVID L. WEBB 188

TELEGRAPHIC REVIEWS 191

COMMENTS

The only way to get rid of responsibilities is to discharge them.
—Walter S. Robertson

Dear Professor X:

I was disappointed that you refused to write a letter for our candidate's tenure case. I was even more disappointed because you waited to tell me until *after* the deadline.

You said you were too busy to write. I *know* you are busy this summer. That's why I wrote to you more than two months in advance to ask whether you would write. I sent you the Curriculum Vitae along with detailed instructions about the nature of the letter and the date it was due. I promised to send you copies of the candidate's papers as well. When you agreed (after a telephone call), I sent those papers promptly. That's proper etiquette for such requests—ample time and ample information.

That was *my* responsibility. Now here is *yours*: When you agree to write, you should do so. Your letter ought to contain enough information to make it clear you have read and judged at least a small portion of the candidate's work. It should contain a short opening paragraph that sets the stage for your specific comments, and it should end with one or two summary sentences that help to guide the reader (and give people preparing the case some quotable sentences). Such letters do not have to be long, nor do they have to contain exquisite detail. They should contain honest opinion, based either on your knowledge of the candidate's written research or your personal contact. And the "letter" needs to be a *signed* letter . . . not an email message or a paragraph dictated to my secretary.

Are there too many requests for such letters? Absolutely—and I'm doing my best to persuade deans and college committees that fewer letters will serve the purpose just as well. But *some* letters will always be necessary in a profession with a system of tenure and promotion like ours.

Writing letters has been a part of our profession for most of this century. Quite likely, a half dozen or so mathematicians took time to write such letters for you in the past. Quite likely, they were busy people as well.

Our profession has been infected by the me-generation philosophy of the 1980's. Writing letters of recommendation, refereeing papers, reviewing grant proposals—for many these are activities without personal profit, and hence (according to the philosophy) without reward. You are not alone; increasing numbers of mathematicians view professional service as the sort of activity other (less busy) people ought to do. Requests for recommendations and referee reports often draw no response—or worse, no results. When most people subscribe to this philosophy, our profession will face a crisis.

I will make phone calls and send email to solicit another letter. It will be hard to find someone else to write, especially since I will have to ask that person to write in a matter of weeks or even days. I *will* find someone, however. And you might find it surprising that the person who finally *does* write the letter turns out to be . . . just as busy as you.

—John Ewing

Yueh-Gin Gung and Dr. Charles Y. Hu Award for Distinguished Service to J. Sutherland Frame

David W. Ballew

“Sud” Frame’s primary interest has always been in his students, their professional growth and their eventual success. He has been instrumental in the growth of Pi Mu Epsilon, personally installing more than fifty Chapters, and in creating and developing in 1952 the highly successful Pi Mu Epsilon Summer Student Paper Conferences in conjunction with the American Mathematical Society and the Mathematical Association of America. He is widely respected as “Dr. Pi Mu Epsilon.” Professor Frame faithfully attends the student presentations, raising questions and making personal comments to most of the young members on how to improve their papers and on the future directions their research might follow.

These innovative efforts spanning more than half a century have given the professional mathematics community a new understanding and appreciation of the research capabilities of undergraduate and Master’s level students. After Dr. Frame established the first National Pi Mu Epsilon Student Paper Conference at Michigan State University in the early 50’s, there came a recognition that more should be done to encourage our best undergraduates to seek careers in research and teaching. In the 1970’s and 1980’s many MAA sections followed his lead and initiated student paper sessions, several in joint sponsorship with Pi Mu Epsilon. The success of student presentations caught attention at the national level and led to the establishment of MAA Student Chapters and additional student paper sessions at national meetings.

James Sutherland Frame earned his Doctorate (1933), Master’s (1930), and Bachelor’s (1929) at Harvard; his main research interest is the Theory of Representation of Finite Groups, a field where he has published over forty of his one hundred and seven papers. He has taught at Harvard, Brown, Allegheny College, and Michigan State University and has appointments at the Institute for Advanced Study and as a Consultant for Graduate Mathematics Programs in Thailand. He has served at the local, regional, national and international levels holding positions on the Board of Governors of the MAA and as Chair or member of many, many scholarly and civic organizations ranging from the Presidency of the Michigan Academy of Arts and Letters to membership on The East Lansing Board of Education and The National Council of the AAUP. He was President of Pi Mu Epsilon for nine years (1957–66) and served as secretary during the organization’s period of most rapid growth (1951–54). He was instrumental in the founding of the Pi Mu Epsilon Journal in 1949 and the first Student Paper Conference in 1952. His many Honors include Phi Beta Kappa, Who’s Who in America and the Senior Research Award of Sigma Xi at Michigan State.

Almost every summer, “Sud” attends as many of the student paper presentations as possible. He immediately grasps the student’s message and, afterwards, speaks to each student in a non-intimidating way to help the student gain deeper insight to improve his or her results. The students are amazed that this grand old gentleman, at 86, can see to the core of the problem so quickly and then see further and clearer than they, who worked so long and so hard. As one student told me after the 1992 Summer Conference, “That man really loves students, doesn’t he?” At the bottom, that says it all. J. Sutherland Frame simply loves students and has spent his life proving that. May he do so for another 86 years!

*Department of Computer Science
Western Illinois University
Macomb, IL 61455*

Counting

*God made the integers, all else is man’s
invention, said Kronecker, but I
prefer to follow Genesis
where God made one and just one
more. Dissatisfied with
only one and two,
Adam and Eve
began the
chain to
more.
One
added
forever,
joined by zero,
matched by opposites,
constructs the integers,
inspires invention of more
new numbers from old—ratios,
radical roots and transcendentals,
transfinite cardinals. Man’s mind is bold.*

*—JoAnne S. Growney
Department of Mathematics and
Computer Science
Bloomsburg University, Bloomsburg, PA 17815*

A New Look at Euler's Theorem for Polyhedra

Branko Grünbaum and G. C. Shephard

1. INTRODUCTION. Euler's Theorem for Polyhedra is one of the most beautiful results of elementary geometry. If v , e and f are, respectively, the number of vertices, edges and faces of a polyhedron P , then the relation

$$v - e + f = 2 \quad (1)$$

is true for cubes, pyramids, prisms, octahedra, and many other polyhedra. One might be tempted to think (as Euler himself apparently did) that this equality holds for *all* polyhedra, but it is easily seen that it fails for the *picture frame* of FIGURE 1(a). Here $v = 16$, $e = 32$ and $f = 16$ so $v - e + f = 0$. The discrepancy is usually dealt with by saying that (1) holds only for polyhedra without any "holes", and then rewriting it in the form

$$v - e + f = 2 - 2g \quad (2)$$

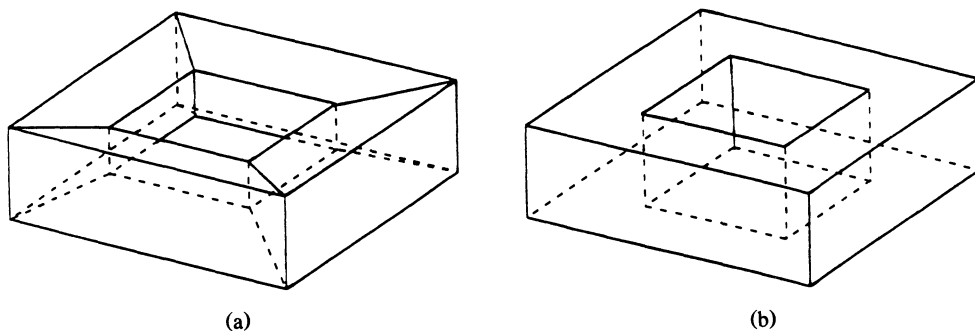


Figure 1. (a) A polyhedron to which Euler's theorem in its elementary form $v - e + f = 2$ does not apply. (b) A polyhedron of genus $g = 1$ to which Euler's theorem in the form $v - e + f = 2 - 2g$ does not apply.

for a polyhedron of "genus" g (that is, "with g holes passing through it", or "with g handles"). Note that here by "polyhedron" we mean the 2-dimensional manifold P which is the boundary of the "solid polyhedron". The quantity on the right-hand side of (2) is usually called the Euler characteristic of that manifold and denoted by $\chi(P)$, so that (2) can be restated as

$$v - e + f = \chi(P). \quad (3)$$

Equations (1), (2) and (3) relate the numbers of vertices, edges and faces of the polyhedron to the topological properties of the polyhedron itself. As a picture frame has just one hole (that is, $g = 1$), relation (2) holds for the polyhedron of Figure 1(a). However, this simple solution is not applicable in all cases. The

polyhedron shown in FIGURE 1(b) also has $g = 1$, but $v = 16$, $e = 24$ and $f = 10$, so $v - e + f = 2$; hence relations (2) and (3) are no longer true. Also, how should one deal with a polyhedron like the funnel-shaped one shown in FIGURE 2 (the boundary of a cube with a pyramid attached to its base and with a pyramidal cavity drilled into it until the apex of the cavity just meets the apex of the attached pyramid)? Does this have a “hole” or not? In this case no integer g fits equation (2)! How should the right side of (1) be modified to deal with this situation, or with the polyhedra of Figure 3? What appeared at first sight to be a simple numerical identity is now seen to be hedged with additional conditions or exceptional cases.

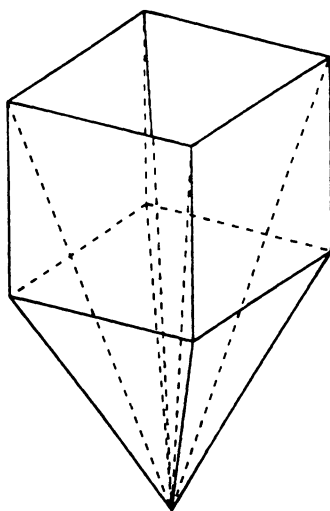


Figure 2. A funnel-shaped polyhedron for which the elementary forms of Euler’s theorem are not valid.

The present paper gives the unexpectedly simple answers to these and related questions. We begin by defining a large class of sets called “polyhedral sets” which generalize familiar polyhedra; they may be closed, open, or neither, connected or not, bounded or not, and their parts may have different dimensions. Examples are shown in FIGURES 1, 2, 3 and later diagrams. For each such polyhedral set P we define an integer $\chi(P)$ called the Euler characteristic of P , and show how this is related to the geometric features of P . While this approach to the Euler characteristic is not new (see references given in Section 6), it has the advantage of allowing very easy determination of the Euler characteristic even for complicated and unusual polyhedral sets. However, although—in the unmatched words of Richard Guy “this is well known to those who well know it”—the circle of those who know it seems to be very small.

To obtain an analogue for polyhedral sets of relation (3) we shall define subsets of P called k -scaffolds (for $k = 0, 1, 2, 3$) and use these to calculate integers V , E , F and C which, in simple cases, correspond to numbers of vertices, edges, faces and cells of P . We then show that

$$V - E + F - C = \chi(P) \quad (4)$$

for *every* polyhedral set P . Clearly, relation (4) is a true generalization of (1), (2) and (3) to polyhedral sets.

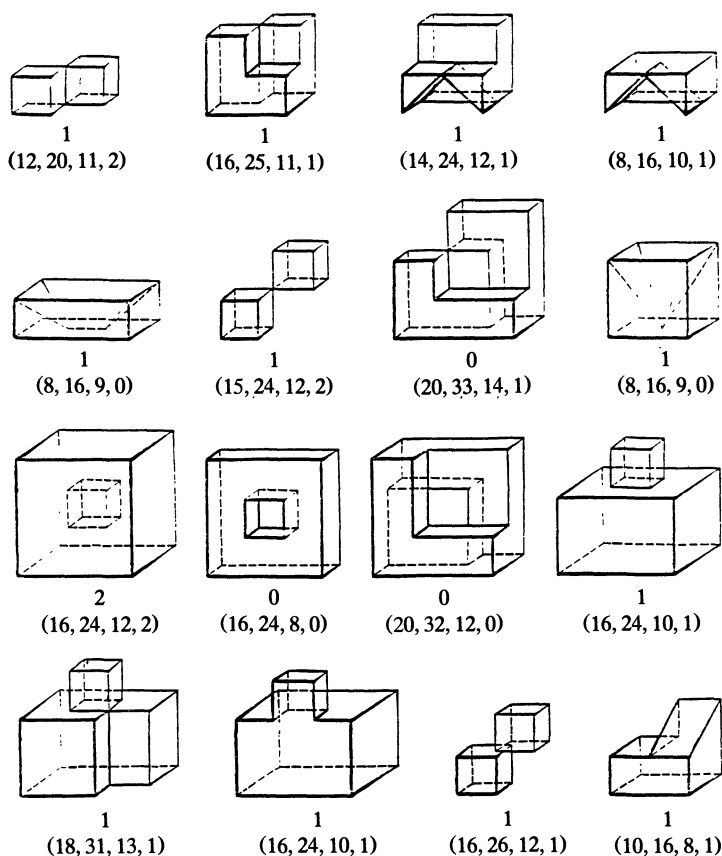


Figure 3. Examples of somewhat unusual polyhedra, after Hajós [8]; we interpret all these solids as closed sets. The single integer in the first line of the caption to each part is the Euler characteristic $\chi(P)$ of the solid P ; it can be determined by the methods of Section 3. The second line of each caption lists (V, E, F, C) ; the meaning of these numbers will be explained in Section 4. The Euler characteristic of the boundary of each polyhedron P is $C + \chi(P)$.

The present approach is simple and elementary; for clarity we shall restrict attention to sets in three-dimensional Euclidean space E^3 , though generalizations to higher dimensions present no difficulties. As defined below, the Euler characteristic is fully additive, and may be positive, negative or zero. Sums (rather than alternating sums), are used in its definition. Notice also that we regard the faces of polyhedral sets as relatively open, in contrast to the traditional approach in which they are considered to be closed sets.

The paper is organized as follows. In Section 2, after some preliminary definitions which may already be familiar to the reader, and are included here in order to avoid any ambiguities, we define polyhedral sets and their dissections. In Section 3 we define the Euler characteristic and establish its fundamental properties. In Section 4 we define the k -scaffolds, establish the Euler relation in the form (4), and present the definition of j -faces. Extensions of these results to unbounded polyhedral sets are presented in Section 5. Section 6 includes a short synopsis of the historical development of the Euler characteristic and related concepts, as well as references to the literature.

2. DEFINITIONS AND TERMINOLOGY. As usual, we shall use $N(x, \delta)$ for the δ -neighborhood of a point x in E^3 . A set is *bounded* if it is contained in some neighborhood of a point. The *interior* of a set S is denoted by $\text{int } S$, and the *boundary* of S by $\text{bd } S$. A set S is *closed* if it contains $\text{bd } S$ and is *open* if it contains no point of $\text{bd } S$, that is, if $S = \text{int } S$.

By a *flat*, or *affine subspace*, we mean any translate of a linear subspace; if the dimension k of a flat is specified, we shall say that it is a k -flat. For any S , we denote by $\text{aff } S$ the *affine hull* of S , that is, the smallest flat that contains S . A set S is said to be k -dimensional if $\text{aff } S$ is k -dimensional. Hence a single point is 0-dimensional, a line segment is 1-dimensional, etc. The dimension of a flat L is denoted by $\dim L$.

The above definitions of interior and boundary apply to sets of any dimension, but for our purposes, *relative* properties are more important. If E is a k -flat and if $N(x, \delta)$ is a neighborhood of a point $x \in E$, then the intersection $N(x, \delta) \cap E$ is called a k -neighborhood of x . Thus a 2-neighborhood of a point x is a small open circular disk of radius δ centered at x , and a 1-neighborhood of x is an *open* line segment (that is, a segment without its endpoints) of length 2δ centered at x . The 0-neighborhood of a point x is just the point x itself.

The *relative interior* of S is the set of those points $x \in S$ for which there exists a k -neighborhood of x , that is contained in S , but for no point y is any $(k + 1)$ -neighborhood of y contained entirely in S ; clearly, there is a unique value of k with this property. Sometimes the relative interior of S will be called its k -interior. The 3-interior is the same as the interior for any set in E^3 . If the k -interior of a set

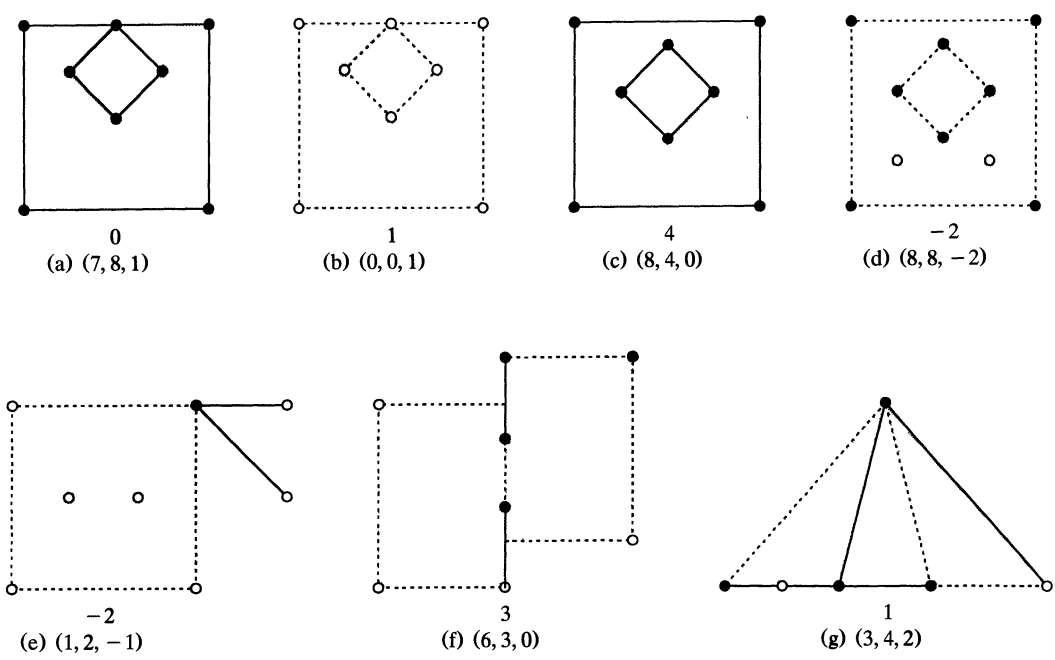


Figure 4. Examples of bounded polyhedral sets P . For simplicity of description and graphical representation, these sets are all in the plane, and all polygonal regions and segments are taken as relatively open. The polygonal regions are shaded. Solid lines and solid dots indicate segments and points included in the set, dashed lines and hollow dots indicate segments and points that are not included. The data in the caption to each part are as in Figure 3, except that C is not listed since it is 0 in all cases. The Euler characteristic of $P \cap \text{relbd } P$ is $\chi(P) - F$.

S is nonempty, we define the *relative boundary* of S as the set of all points of $\text{bd } S$ which do not belong to the relative interior of S . The relative interior and relative boundary of S are denoted by $\text{relint } S$ and $\text{relbd } S$, respectively. The set S is said to be *relatively open* if $S = \text{relint } S$.

A single point, an open ray or segment, a straight line, a plane, or the whole space E^3 are examples of relatively open sets. If Q is a square region in E^3 (so $\text{aff } Q$ is a plane, and Q is 2-dimensional) then $\text{relbd } Q$ consists of the union of four open segments (each of which is the relative interior of one of the sides of the square) together with the four vertices of the square. Further, $\text{relint } Q$ is the part of $\text{aff } Q$ that lies inside $\text{relbd } Q$. If S is the boundary of a cube then S contains no points whose 3-neighborhoods lie in S , hence $\text{int } S$ is the empty set. The points whose 2-neighborhoods lie in S are the points in the faces of the cube, that is, the points of S apart from the edges and vertices. This set is therefore the 2-interior of S . Moreover, since $\text{int } S$ is empty, the 2-interior of S is, by definition, the relative interior of S . Thus $\text{relint } S$ is the union of the relative interiors of the six square

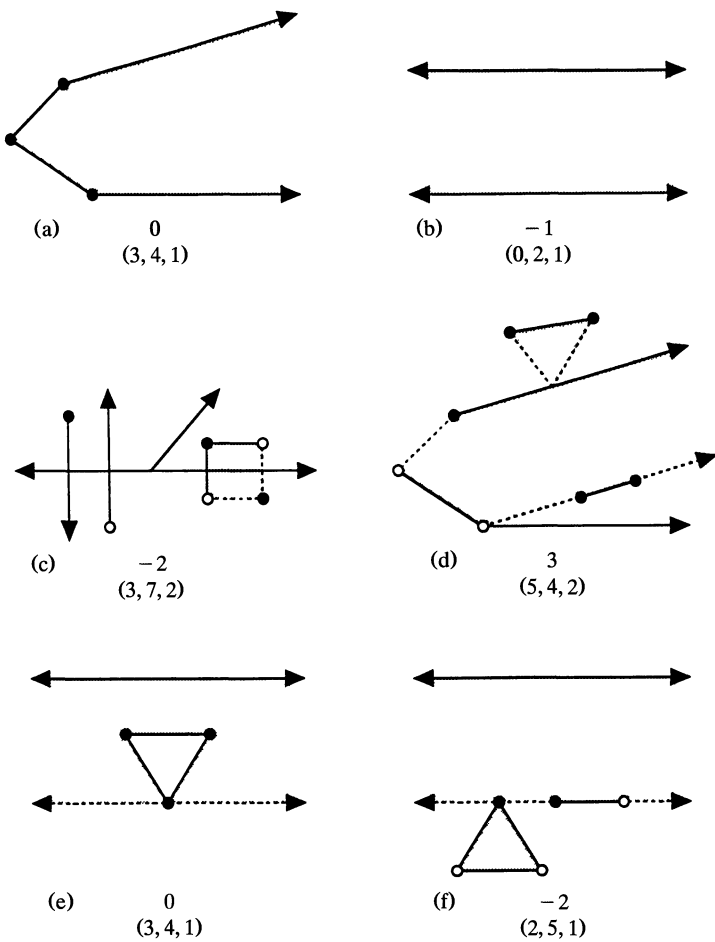


Figure 5. Examples of unbounded polyhedral sets in the plane. The conventions are the same as in Figure 4; in addition, arrowheads are used to distinguish halflines and lines from segments. (a) shows a line-free closed convex set, and the set in (b) is closed and convex, but not line-free (L is a line and $k = 1$ in the notation of Section 5).

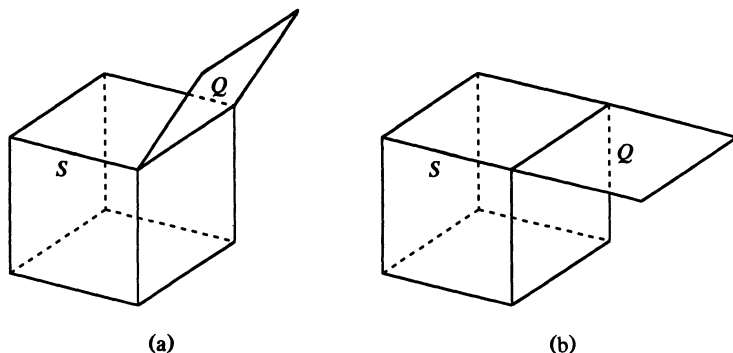


Figure 6. An illustration of the dependence of the relative interior of a set on the mutual position of its parts. In (a), the set $T = S \cup Q$, where S is the boundary of a cube and Q is a closed square not coplanar with any face of the cube. Hence $\text{relint } T = \text{relint } S \cup \text{relint } Q$. In (b), for the set $U = S \cup Q$, in which Q is coplanar with the top face of the cube, meeting it along the edge E , we have $\text{relint } U = \text{relint } S \cup \text{relint } Q \cup \text{relint } E$. The Euler characteristic is 2 in both cases, as can be verified using the definition given in Section 3. In the notation introduced in Section 4, $\text{scaf}_2 T$ consists of seven open squares, $\text{scaf}_1 T$ consists of 15 open segments, and $\text{scaf}_0 T$ of 10 vertices; hence $V = 10$, $E = 15$, and $F = 7$. On the other hand, $\text{scaf}_2 U$ consists of five open squares and one open rectangle, $\text{scaf}_1 U$ of 12 open segments, and $\text{scaf}_0 U$ of eight points.

faces of the cube. If T consists of the set S just considered together with a square Q which is not coplanar with any face of the cube (see FIGURE 6(a)) attached to S along a common edge E , then $\text{relint } T = \text{relint } S \cup \text{relint } Q$; but for a set U (see FIGURE 6(b)) in which Q is coplanar with a face of S we have $\text{relint } U = \text{relint } S \cup \text{relint } Q \cup \text{relint } E$. A more complicated example is shown in FIGURE 7 and explained in the caption.

A set S is called *convex* if, given any two distinct points $x, y \in S$, the closed line segment with endpoints x and y lies entirely in S . Thus a line segment is necessarily convex; a single point and the empty set are also convex, since the definition is vacuous in this case. A *closed half-space* is the (unbounded convex) set of points that lie to one side of, or on, a plane in E^3 . Any set which is the intersection of a finite number of closed half-spaces is called a *closed convex polyhedron*. Familiar examples of closed convex polyhedra are (closed) cubes, (closed) squares, (closed) line segments, and single points. These are of three, two, one and zero dimensions respectively. But our definition includes also unbounded convex polyhedra, such as the examples in FIGURES 5(a) and 5(b); their edges can be rays (halflines) or straight lines, and faces and cells can be unbounded as well.

Euler's Theorem in its basic form (1) holds for the boundary of 3-dimensional closed and bounded convex polyhedra. Many elementary proofs of this fact are known, see Section 6. The case of unbounded polyhedra will be considered in Section 5.

A *relatively open convex polyhedron* is the relative interior of a closed convex polyhedron. It should be noted that whereas an open convex polyhedron (which is necessarily 3-dimensional) is the intersection of a finite number of *open* half-spaces, the same is *not* true for relatively open convex polyhedra of dimension less than three. On the other hand, if P is a relatively open convex polyhedron lying in a d -flat $E = \text{aff } P$ (where $d = 1$ or 2) then P can be written as the intersection of finitely many open half d -flats (each of which is a relatively open set) lying in E . Alternatively, P is the intersection of E with a family of open halfspaces. As we

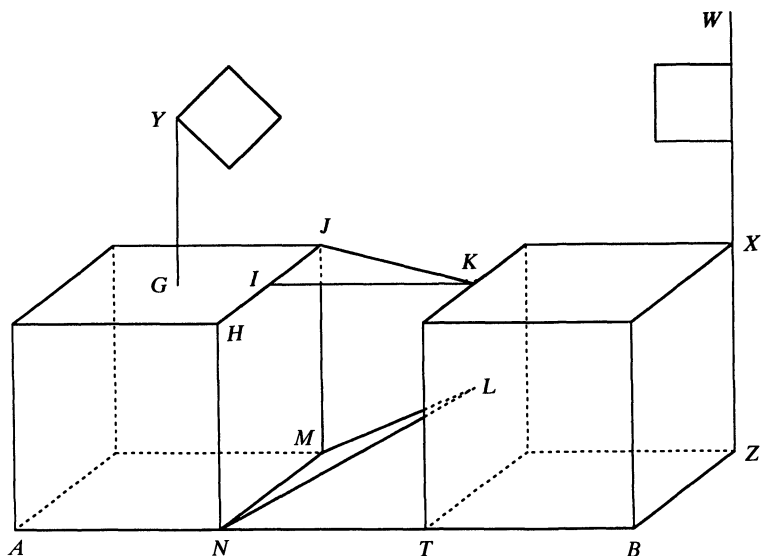


Figure 7. A polyhedral set P (assumed closed, to simplify the discussion) consisting of two solid cubes joined by two triangles and a segment, and of two squares attached to segments which touch the cubes; clearly $\text{int } P = \text{scaf}_3 P$ consists of the two open cubes, so $C = 2$. The polyhedral set $Q = \text{bd } P$ consists of the 14 squares shown, together with the two triangles and three segments. Its relative interior $\text{relint } Q = \text{scaf}_2 Q = \text{scaf}_2 P$ consists of the relative interiors of the squares and the triangles, together with the open segment IJ along which the upper triangle is joined to the cube; hence $F = 15$. The set $R = \text{relbd } Q$ consists of the following open segments: all the edges of the two cubes except HJ , the four edges of one of the squares and three of the other, and the segments $GY, HI, IK, JK, LM, LN, NT, WX$; moreover, R also contains all the vertices of the cubes and squares, and the point W (but not the points G and L). It follows that $\text{relint } R = \text{scaf}_1 R = \text{scaf}_1 Q = \text{scaf}_1 P$ consists of the following open segments: AB, GY, HI, WZ , twenty edges of the two cubes, seven edges of the two squares, and four edges of the two triangles; hence $E = 35$. Finally, $\text{relbd } R = \text{scaf}_0 R = \text{scaf}_0 Q = \text{scaf}_0 P$ consists of all the vertices of the two cubes except N, T, X , as well as of the four vertices of one of the squares and two of the other, and the points I and W ; hence $V = 21$. It follows that $\chi(R) = -1$, $\chi(Q) = 1$, and $\chi(P) = -1$. These values can easily be verified by directly using the definition of the Euler characteristic.

shall see, relatively open convex polyhedra play a central rôle in our treatment of the Euler characteristic.

From the definition it is clear that the intersection of any two closed convex polyhedra is a closed convex polyhedron. Analogously, though this requires some additional reasoning, it may be verified that regardless of their dimensions the intersection of any two relatively open convex polyhedra is a relatively open convex polyhedron. Following convention, we shall sometimes refer to a bounded convex polyhedron of two dimensions as a *polygon*, and a bounded convex polyhedron of one dimension as a *line segment* or simply *segment*; in each case we must, of course, specify whether the polyhedron is closed, relatively open, or neither.

If a set P is the union of members a finite family $\mathcal{C} = \{C_1, C_2, \dots, C_n\}$ of pairwise disjoint sets C_1, C_2, \dots, C_n , we say that \mathcal{C} is a *dissection* of P and we write $P = \bigcup \mathcal{C}$, where the dot in the union symbol indicates that the sets C_i are pairwise disjoint. If P is a set that admits a dissection \mathcal{C} all elements of which are relatively open convex polyhedra we shall say that P is a *polyhedral set*, and express this by writing

$$P = \bigcup_c \mathcal{C} \quad (5)$$

where the subscript c is to remind us that each element of \mathcal{C} is a relatively open convex polyhedron. Such a family \mathcal{C} is called a *relatively open convex dissection* of P and each element of \mathcal{C} is called an *element* of the relatively open convex dissection. All the sets shown in FIGURES 1 to 9 are polyhedral sets in this sense. It will be observed that the definition is very general in that it does not require P to be open or closed, connected, simply connected or even *homogeneous* in the sense that neighborhoods of all points of P are of the same dimension. Nor are there any restrictions on the incidences of the closures of the elements. On the other hand, a bounded and closed polyhedral set is necessarily compact. The collection of all polyhedral sets will be denoted by \mathbb{P} . These polyhedral sets are the sets to which relation (4) will apply.

It is clear that except when P consists of a finite number of points, its expression (5) as a disjoint union of relatively open convex polyhedra is not unique. In fact, every relatively open convex polyhedron of dimension greater than 0 admits infinitely many open dissections.

Let $P = \bigcup_c \mathcal{C}$ and $P = \bigcup_c \mathcal{F}$ be two dissections of P ; we shall say that the latter is a *refinement* of the former if each element of \mathcal{C} is a disjoint union of elements of \mathcal{F} . If we are given *any* two dissections $P = \bigcup_c \mathcal{C}$ and $P = \bigcup_c \mathcal{D}$ of P then it is possible to find another dissection $P = \bigcup_c \mathcal{F}$ which is a refinement of each; any such \mathcal{F} will be called a *common refinement* of the dissections \mathcal{C} and \mathcal{D} . In fact it suffices to take, as elements of \mathcal{F} , all non-empty sets of the form $C \cap D$,

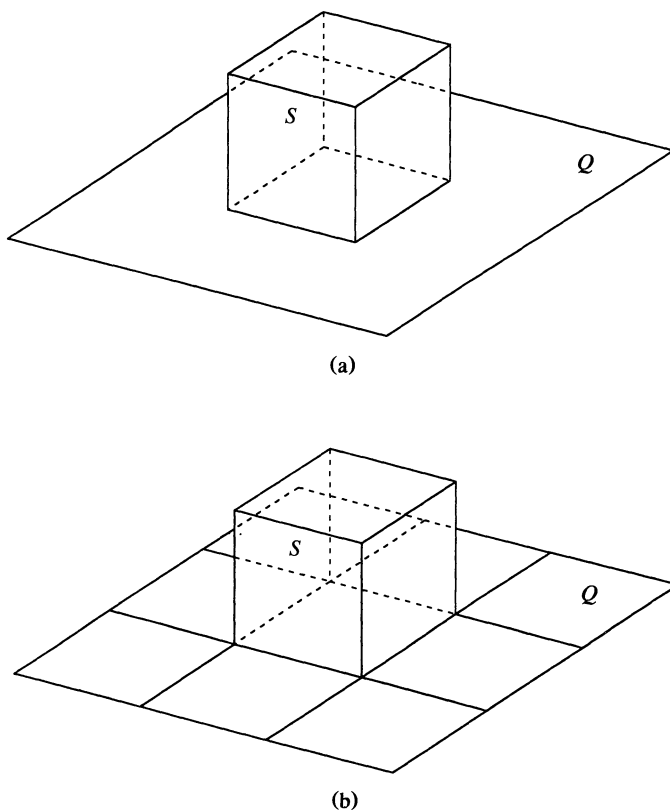


Figure 8. (a) A polyhedral set P consisting of an open cube S and a relatively open square Q . This dissection of P is not complex-like. (b) A relatively open convex dissection of P which is complex-like.

where $C \in \mathcal{C}$, and $D \in \mathcal{D}$. From the properties of relatively open convex sets it is immediate that if the dissections \mathcal{C} and \mathcal{D} are relatively open convex dissections, then the common refinement \mathcal{F} is another dissection of the same kind. Moreover, among such common refinements of the relatively open convex dissections \mathcal{C} and \mathcal{D} of a polyhedral set P it is possible to find a common refinement \mathcal{F} which is *complex-like*. By this we mean that whenever an element of \mathcal{F} meets the relative boundary of another element of \mathcal{F} , it is contained in that boundary, see FIGURE 8.

These simple observations are basic for all that follows.

3. THE EULER CHARACTERISTIC. Throughout this section we shall restrict attention to bounded polyhedral sets; the extension of these results to unbounded sets will be given in Section 5.

For any polyhedral set $P \in \mathbb{P}$ we define an integer $\chi(P)$ in the following way:

- (a) $\chi(\emptyset) = 0$.
- (b) If P is a relatively open convex polyhedron of dimension d , then $\chi(P) = (-1)^d$.
- (c) If $P = \bigcup_{C \in \mathcal{C}} C$ is a relatively open convex dissection of P then $\chi(P) = \sum_{C \in \mathcal{C}} \chi(C)$.

This integer $\chi(P)$ is called the *Euler characteristic* of P . Part (c) of the definition appears to imply that the definition of $\chi(P)$ depends on the dissection \mathcal{C} that is used. However, this dependence is only apparent. This, and other important properties of $\chi(P)$ are given by the following theorems:

Theorem 1. *The Euler characteristic $\chi(P)$ is well-defined in the sense that its value, as given by (c) above is independent of the relatively open convex dissection \mathcal{C} of P used in the computation.*

Theorem 2. *If P is a closed and bounded convex polyhedron then $\chi(P) = 1$.*

Theorem 3. *The Euler characteristic $\chi(P)$ satisfies the valuation property: if $P_1 \in \mathbb{P}$ and $P_2 \in \mathbb{P}$, then*

$$\chi(P_1) + \chi(P_2) = \chi(P_1 \cap P_2) + \chi(P_1 \cup P_2). \quad (6)$$

It should be noted that Theorem 2 refers to the Euler characteristic of the (bounded and closed) convex polyhedron itself, and not—as in the discussion in Section 1—to the Euler characteristic of its boundary.

In some treatments of the Euler characteristic Theorems 2 and 3 are used to define $\chi(P)$ for closed convex sets and for those sets which can be obtained from them by taking finite unions. Thus the definition of $\chi(P)$ given here may be regarded as an extension of the traditional approach to sets which need not be closed. On the other hand, our definition is restricted to polyhedral sets, hence not applicable to non-polyhedral sets even if they are convex.

The proofs of Theorems 1, 2 and 3 are omitted since they follow the usual techniques. We note only that for Theorem 1 we rely on the fact mentioned earlier that two convex dissections \mathcal{C} and \mathcal{D} of a polyhedral set P have a complex-like common refinement \mathcal{F} .

Immediate consequences of these theorems include the following:

Corollary 1. *If \mathcal{C} is any dissection of the polyhedral set P such that each $C \in \mathcal{C}$ is a polyhedral set, then*

$$\chi(P) = \sum_{C \in \mathcal{C}} \chi(C).$$

Corollary 2. (Inclusion-exclusion property). *If \mathcal{C} is any family of polyhedral sets such that $P = \bigcup \mathcal{C}$, then*

$$\chi(P) = \sum \chi(C) - \sum \chi(C_1 \cap C_2) + \sum \chi(C_1 \cap C_2 \cap C_3) - \cdots$$

where the first sum is over all $C \in \mathcal{C}$, the second sum is over all sets $C_1, C_2 \in \mathcal{C}$ of two distinct elements of \mathcal{C} , the third sum is over all sets of three distinct elements of \mathcal{C} , etc.

Corollary 3. *For any polyhedral set $P \in \mathbb{P}$,*

$$\chi(P \cap \text{relbd } P) = \chi(P) - \chi(\text{relint } P);$$

hence, if P is closed, $\chi(\text{relbd } P) = \chi(P) - \chi(\text{relint } P)$.

Using the above results and suitable dissections, it is easy to verify that each of the polyhedral sets in FIGURES 3, 4 and 5 has the Euler characteristic indicated.

4. EULER'S THEOREM FOR BOUNDED POLYHEDRAL SETS. We now show how the Euler characteristic, as introduced in Section 3, can be used to derive analogues of the traditional Euler equations (1) and (2) which are valid for bounded but otherwise general polyhedral sets. The basic approach is to express each such polyhedral set $P \in \mathbb{P}$ as the disjoint union of four well-determined sets, called *k-scaffolds* of P and denoted $\text{scaf}_k P$ for $k = 0, 1, 2, 3$.

The 3-scaffold $\text{scaf}_3 P$ is simply $\text{int } P$, the set of all interior points of P . We delete $\text{scaf}_3 P$ from P , leaving $P_2 = P \setminus \text{scaf}_3 P$; clearly, $P_2 \subset \text{bd } P$. The 2-scaffold $\text{scaf}_2 P$ is the set of all 2-interior points of the set P_2 . (Equivalently, $\text{scaf}_2 P$ is the relative interior of the set $P \cap \text{bd } P$.) Similarly, $\text{scaf}_1 P$ is the set of all 1-interior points of $P_1 = P_2 \setminus \text{scaf}_2 P$, and $\text{scaf}_0 P$ is the finite set of points $P_0 = P_1 \setminus \text{scaf}_1 P$.

Each *k-scaffold* of P is a relatively open set; in fact, it is a relatively open polyhedral set, and $S = \{\text{scaf}_0 P, \text{scaf}_1 P, \text{scaf}_2 P, \text{scaf}_3 P\}$ is a relatively open dissection of P into polyhedral sets. To verify this claim, we only have to show that each $\text{scaf}_k P$ is a polyhedral set. This follows at once from the observations:

(i) the definition of $\text{scaf}_k P$ is independent on the relatively open convex dissection \mathcal{C} which establishes that P is a polyhedral set, and

(ii) each relatively open convex dissection of P can be refined to one that is complex-like. For such a refinement it is clear that each element is contained in one and only one scaffold of P , and so defines a relatively open convex dissection of each $\text{scaf}_k P$.

As a consequence we have:

Corollary 4. *If p is a bounded polyhedral set then*

$$\chi(P) = \chi(\text{scaf}_0 P) + \chi(\text{scaf}_1 P) + \chi(\text{scaf}_2 P) + \chi(\text{scaf}_3 P).$$

We note that if P is a convex polyhedron, then $\text{scaf}_2 P$ is the union of the relative interiors of the faces of P , $\text{scaf}_1 P$ is the union of the relative interiors of

the edges of P , and $\text{scaf}_0 P$ is the set of vertices of P . These remarks motivate the following notation: $V = \chi(\text{scaf}_0 P)$, $E = -\chi(\text{scaf}_1 P)$, $F = \chi(\text{scaf}_2 P)$, and $C = -\chi(\text{scaf}_3 P)$. Note that we use upper case symbols in order to stress the distinction between the precisely defined entities and the somewhat vague quantities mentioned in the Introduction and Section 1. Corollary 4 implies:

Theorem 4. *If P is a polyhedral set then*

$$V - E + F - C = \chi(P).$$

Thus relation (4) has been established.

Theorem 4 will now be illustrated by means of examples.

Consider first the closed polyhedral set P of FIGURE 9(a); it is to be understood as a solid, the interior of which is $\text{scaf}_3 P$. The 2-scaffold $\text{scaf}_2 P$ consists of 12

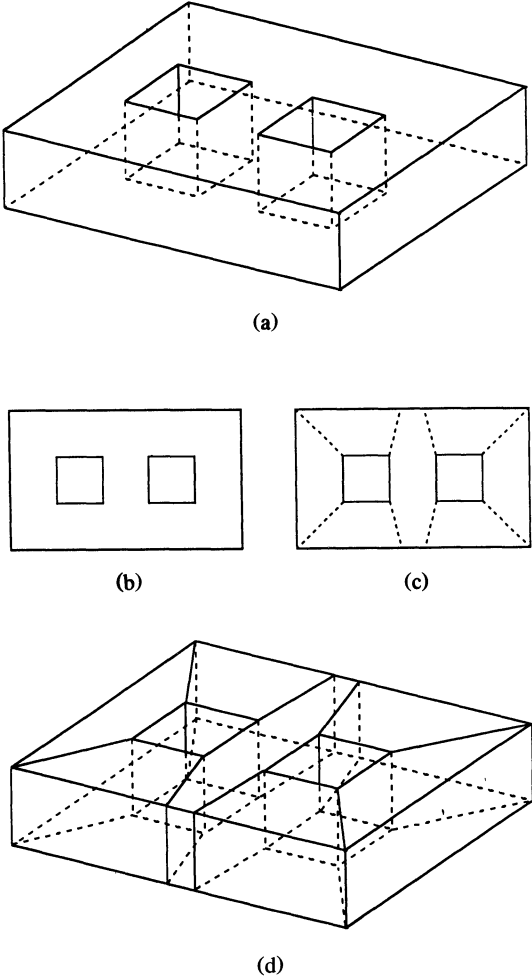


Figure 9. (a) A (closed) polyhedral set; if interpreted as a solid P , it has $V = 24$, $E = 36$, $F = 10$, $C = -1$, and $\chi(P) = -1$. If P is interpreted as a 2-manifold then $\chi(P) = -2$. (b) One of the polygons B which occurs as the upper or lower face of the polyhedral set P shown in (a), and (c) a dissection of B . Interpreting B as a relatively open set, this shows that $\chi(B) = -1$. (d) A dissection of $\text{int } P$, which shows that $\chi(\text{int } P) = 1$.

relatively open rectangular faces (each of which has Euler characteristic equal to 1) and two other relatively open polygons, namely the top and bottom of P , each of which is an open rectangle from which two closed squares have been removed. (FIGURE 9(b)). The Euler characteristic of each of these polygons (which we shall denote by B) can be easily determined using the open dissection shown in FIGURE 9(c). This shows B expressed as the union of seven relatively open convex polygons and eight open line segments, so $\chi(B) = 7 \times 1 + 8 \times (-1) = -1$. We deduce that $F = 12 \times 1 + 2 \times (-1) = 10$. As $\text{scaf}_1 P$ consists of 36 open segments, each with Euler characteristic -1 , it follows that $E = 36$. The 24 vertices form $\text{scaf}_0 P$, hence $V = 24$. To determine C , consider the open convex dissection of $\text{int } P = \text{scaf}_3 P$ indicated in FIGURE 9(d); it consists of seven open 3-dimensional convex polyhedra, and of 8 relatively open convex polygons, hence $C = -\chi(\text{scaf}_3 P) = -(7 \times (-1) + 8 \times 1) = -1$. Therefore

$$\chi(P) = V - E + F - C = 24 - 36 + 10 - (-1) = -1.$$

The value of $\chi(P)$ can be verified by using the inclusion-exclusion principle (Corollary 2). We represent P as the union of seven closed convex sets, all of which are prisms (six on quadrangular bases, and one on an octagonal base—FIGURE 9(d) can also be interpreted as showing this representation). The intersections by pairs are eight closed polygons (rectangles), and every set of three has empty intersection. Hence $\chi(P) = 7 - 8 = -1$, as before.

It should be noted that the flexibility built in into the approach followed here enables one in many cases to avoid subdivisions when calculating $\chi(P)$. For the example in FIGURE 9(a) we could argue that P , together with two open rectangular boxes and four relatively open squares, forms a dissection of a closed rectangular box. Therefore $1 = \chi(P) + 2(-1) + 4$, leading to the same value $\chi(P) = -1$.

The numbers calculated above can also serve to illustrate Corollary 3, since

$$\chi(\text{relbd } P) = 24 - 36 + 10 = -2,$$

which coincides with

$$\chi(P) - \chi(\text{relint } P) = -1 - 1 = -2,$$

since $\chi(\text{relint } P) = -C = 1$.

As a second example, consider the closed polyhedral set P in FIGURE 2. Using a plane through two opposite vertical edges of the cube, partition P into two parts, each of which is a closed polyhedron and has Euler characteristic equal to 1. The intersection of these parts is the union of two closed triangles with a vertex in common, and again this has Euler characteristic 1. Hence, by the inclusion-exclusion principle, $\chi(P) = 2 - 1 = 1$. Considering scaffolds, we can verify $V = 9$, $E = 20$, $F = 12$, and so, from Corollary 4,

$$C = -\chi(\text{scaf}_3 P) = -\chi(\text{int } P) = 1 - 12 + 20 + 9 = 0.$$

This can be verified using a dissection of $\text{int } P$ analogous to that just used for P . Each half of $\text{int } P$ has Euler characteristic -1 , but their intersection now has characteristic 2, since it consists of two disjoint, relatively open triangles. Thus $\chi(\text{int } P) = 2 \times (-1) + 2 = 0$, as before. For another verification we may consider the dissection consisting of P , an open 4-sided pyramid, and an open square, from which we have $1 = \chi(P) + (-1) + 1$, hence $\chi(P) = 1$.

Additional examples illustrating our theorems and corollaries appear in FIGURE 3. The quantities V, E, F, C are indicated in the form of a vector (V, E, F, C) below each part of the diagram. Each of the polyhedral sets is considered to be solid, that is, to have non-empty interior, and to be equal to the closure of its

interior. Other examples are shown in FIGURES 4, 6, 7, 9 and described in the captions.

Finally, we note that the use of scaffolds leads to a definition of j -faces ($j = 0, 1, 2, 3$) which is applicable to *all* polyhedral sets. We define a j -face of a polyhedral set P to be a connected component of $\text{scaf}_j P$. For each j , these components are well determined. In the case of convex polyhedra—and many others as well—these j -faces are the relative interiors of faces and edges in the traditional approach to the subject. However, in more complicated situations rather unusual sets can appear as faces. For example, in FIGURE 4, one 1-face of the first polyhedral set in (a) consists of the union of three open segments, namely the upper segment and the two contiguous sides of the deleted square. In FIGURE 4(g), one 1-face is the union of two open segments. In FIGURE 7, one 2-face consists of the open segment IJ and the two open squares and one open triangle that contain IJ in their boundary; another 2-face consists of the open square that contains the point L , and the open triangle with vertex L , while one 1-face consists of the open segment AB together with the five additional open segments that have N or T as an endpoint. It remains to be seen whether this generalization of the usual concept of “face” will lead to interesting mathematics.

5. UNBOUNDED POLYHEDRAL SETS. Now we shall extend the results of the previous sections to unbounded polyhedral sets. As we pointed out in Section 2, our definition of a closed polyhedron applies equally to the unbounded case. We may also allow the elements in the definition of a polyhedral set (5) to be unbounded, and thus extend the family \mathbb{P} to include unbounded sets. From now on, \mathbb{P} will be used in this wider sense. We shall assume that the reader has some familiarity with the structure of unbounded convex sets as explained, for example, in Grünbaum [1967], Section 2.5 and the references in Section 2.7.

For any non-empty convex polyhedron P in the d -dimensional space E^d , let L represent a k -flat, contained in P , and chosen to have maximal possible dimension k . Then P is called *line-free* if and only if $k = 0$. A bounded convex polyhedron is necessarily line-free, as is also a set such as that shown in FIGURE 5(a). For the convex polyhedron in FIGURE 5(b), k takes the value 1. It is known (see Grünbaum [1967], page 24) that if P is a closed, convex polyhedron then it may be written as a “direct product” or “vector sum”

$$P = L \oplus \tilde{P}, \quad (7)$$

where L is the maximal k -flat defined above and \tilde{P} is a line-free polyhedron whose dimension is k less than the dimension of P . The sign \oplus means that every point $x \in P$ can be written uniquely in the form $x = y + z$ (vector addition) where $y \in L$ and $z \in \tilde{P}$. We may take for \tilde{P} any set of the form $P \cap \tilde{L}$, where \tilde{L} is a $(d - k)$ -flat orthogonal to L in E^d . We shall refer to (7) as the *standard linear decomposition* of P . In the example of FIGURE 5(b), L is a line and \tilde{P} is a segment.

For the boundary of a 3-dimensional unbounded line-free convex set P a relation analogous to (1) holds, namely

$$v - e + f = 1, \quad (8)$$

where v, e, f are, respectively, the numbers of vertices, edges and faces of P (see Grünbaum [1967], Section 8.5). If P is 2-dimensional, the corresponding result for the relative boundary is

$$v - e = -1. \quad (9)$$

An example is provided by the line-free set in FIGURE 5(a), where $v = 3$ and $e = 4$; two of the edges are line segments, and the other two are rays (unbounded half-lines).

Just as the definition of the family \mathbb{P} in Section 2 was formulated in such a way as to apply also to unbounded sets, so were Theorems 1 and 3, together with the three corollaries in Section 3 phrased in such a way as to apply also in the unbounded case. The result analogous to Theorem 2 is as follows:

Theorem 2*. *Let P be a closed convex polyhedron, and $P = L \oplus \tilde{P}$ be a standard linear decomposition of P with $\dim L = k$, then*

$$\chi(P) = \begin{cases} 0 & \text{if } \tilde{P} \text{ is unbounded} \\ (-1)^k & \text{if } \tilde{P} \text{ is bounded.} \end{cases}$$

Proof of Theorem 2.* The case where P is bounded has been dealt with in Section 3, Theorem 2. Now let P be unbounded but line-free. If P is a closed half-line (ray), then $\chi(P) = 1 - 1 = 0$. If P is 2-dimensional then, by (9),

$$\chi(P) = v - e + 1 = 0,$$

and if P is 3-dimensional then, by (8),

$$\chi(P) = v - e + f - 1 = 0.$$

If P is not line-free, then to each j -dimensional element of P there corresponds a $(j - k)$ -dimensional element of \tilde{P} (note that all elements of P have dimensions $\geq k$). Hence $\chi(P) = (-1)^k \chi(\tilde{P})$ and equals $(-1)^k$ or 0 depending on whether \tilde{P} is bounded or not. This completes the proof of the theorem.

The definitions of the scaffolds of a polyhedral set P , together with Theorem 4 and Corollary 4, hold with trivial modifications in the unbounded case. Examples of the application of Theorem 4 to unbounded sets appear in FIGURE 5. Details are given in the caption to the figure.

6. HISTORICAL REMARKS AND COMMENTS. The history of Euler's Theorem and concepts related to it are both interesting and voluminous. It involves many of the ideas that led to modern algebraic topology, and also many of the errors which were committed in that development. At least two books have been devoted to the early history and attempts at clarification of Euler's Theorem (Lakatos [13], Federico [4]), and countless books and articles contain short accounts. Here only a very brief survey will be given.

Euler first published his theorem in 1750, stating that he had no satisfactory proof but was convinced of its general validity by a wealth of examples. (The frequently encountered assertion that Euler's Theorem was known to Descartes a century before Euler is unsupported by any evidence, and based on erroneous interpretation of some of Descartes' writings. For a discussion of this and other historical errors concerning Euler's Theorem see Malkevitch [17] and the references given there, and Federico [4].) Euler's formulation was, essentially, that "for every solid bounded by flat surfaces, the number of surfaces increased by the number of vertices exceeds by two the number of edges." Later, Euler presented a proof of the theorem, as did several other mathematicians. Early in the nineteenth century it was observed that the assertion cannot be true in the generality claimed

(L'Huilier [16], Hessel [9]). In a masterpiece of understatement, Hessel remarks:

Other excellent mathematicians (Legendre, Cauchy, Gergonne, Rothe and Steiner) supplied proofs for the general validity of the theorem. But in fact, it suffers from exceptions.

In order to illustrate such exceptions and the difficulties in removing them, Hessel shows pairs of polyhedra such that “for each shape for which Euler’s rule is not valid, a rather similar one can be found for which the rule is valid.” Hessel’s examples are reproduced in FIGURE 10; we shall return to them shortly.

These developments led to a number of reformulations of the basic version of Euler’s Theorem (given by relation (1)) through the introduction of various parameters that replaced the value 2 of the right-hand side. All the discussions were dogged by two difficulties.

On the one hand, no precise definitions were given for the polyhedra under consideration or for their faces, edges and vertices. It was more or less generally assumed that one is dealing with solids and considering features on their surfaces—but how to determine faces (or edges) was illustrated by examples rather than defined by unambiguous rules. A glance at the collection of examples of polyhedra in FIGURE 3, taken from Hajós [8], should convince the reader that the concepts of face, edge and vertex are not very straightforward to define even for those special polyhedra which are compact and such that the boundary of the polyhedron coincides with the boundary of its interior. This observation explains why, in the later decades of the nineteenth century, Euler-type relations with more and more complicated right-hand sides were appearing in the literature.

On the other hand, in its original formulation and in the minds of many early workers, it was the hallmark of Euler’s theorem that it involved only the *numbers* of vertices, edges and faces—*without any consideration of the nature of the faces*. Thus, in Hessel’s presentation, his even-numbered examples are “good” because for them $v - e + f = 2$, and the odd-numbered polyhedra are “bad” because the numbers to be inserted in the left-hand side do not yield 2. Remarkably, some of this attitude survived even to our days: in Seydel [30, page 322] the same example as in Hessel’s diagram labelled 2 (and our FIGURE 1(b)) is given, with the comment that this illustrates the validity of relation (1) for *all* polyhedra! We shall return to the faces of polyhedra later in this discussion.

The second half of the nineteenth century saw the gradual clarification of the difficulties; one step was the insight that relations such as (1) deal with the surface, and not directly with the solid. The concept of genus of those special surfaces which are called orientable manifolds helped to reach the relation (2). In particular, this led to the understanding that the original formulation of Euler’s theorem applies to the boundary of 3-dimensional convex polyhedra and, more generally, to maps on the sphere and other closed surfaces.

At the same time, Euler’s theorem was extended to higher dimensions as one of the first results in the emerging discipline of algebraic topology. It was established (or, at least, stated!) that the Euler characteristic equals to the alternating sum of so-called “Betti numbers”. But again, there were more good intentions than mathematically proved results. In the words of Dieudonné [3]:

... the mathematicians of the second half of the nineteenth century which were busy with these questions [of algebraic topology] speak freely of curves, of surfaces, of deformations, ..., without ever saying what they mean by these words.

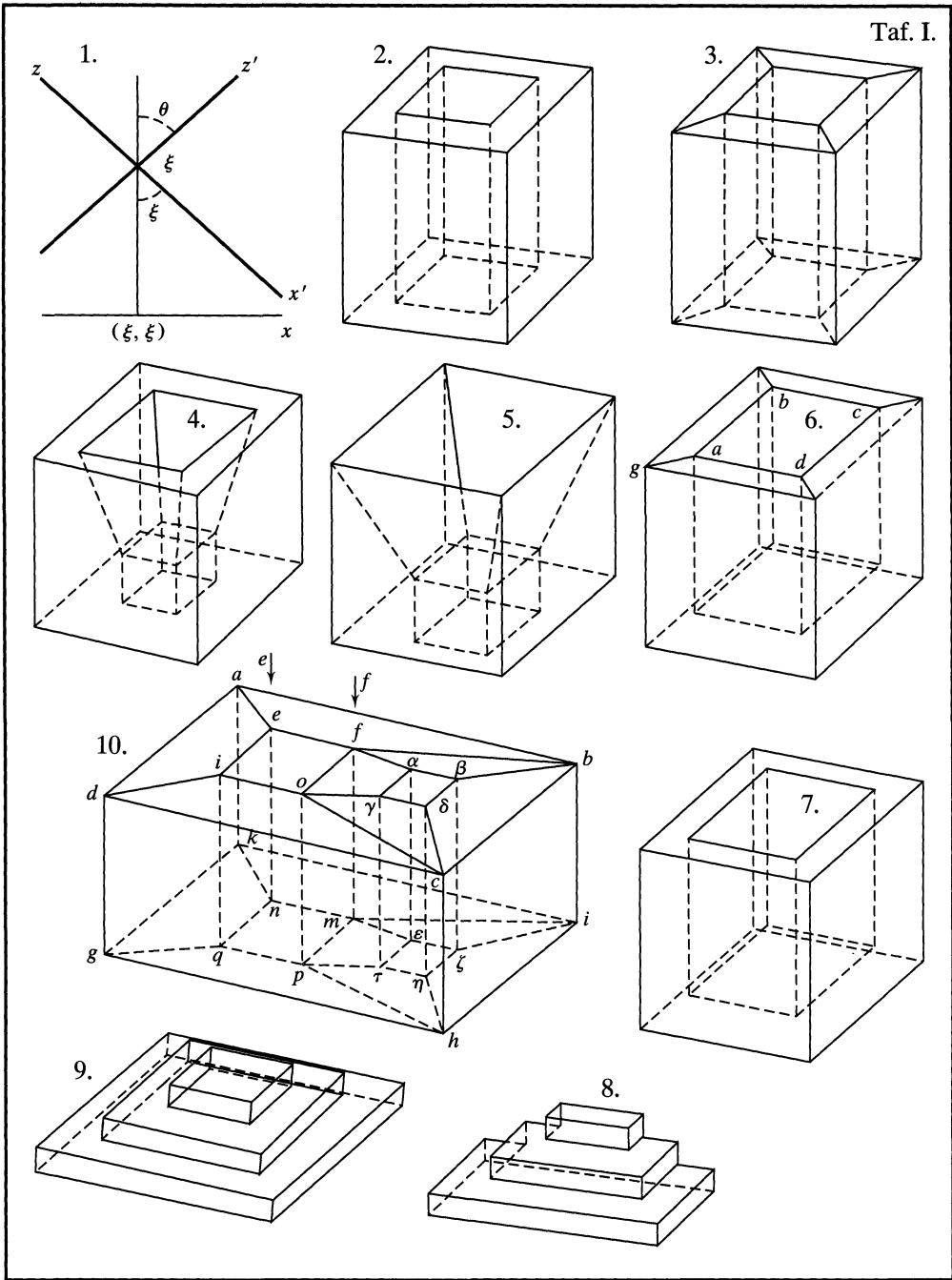


Figure 10. Plate taken from Hessel [9]. These diagrams illustrate the difficulties that early authors had in extending the simple Euler relation (1) to more complicated polyhedra. Parts 2 to 5 show a cube with a hole drilled through it, and parts 6 and 7 show a cube with an indentation. Part 10 shows a block with two prismatic holes through it. In parts 3, 6 and 10 some faces have been subdivided.

Starting with Poincaré [26] and completed, simplified and extended by others (in particular, Alexander [1]) these concepts involved in the definition of Euler characteristic acquired precise meaning for simplicial complexes and for more general objects; see, for example, Hilton-Wylie [11], Singer-Thorpe [32], or any other modern text on algebraic topology. However, this development led to a loss of the connection to the origins of Euler’s theorem as a relation involving the vertices, edges and faces of a polyhedral object. Apart from these developments, in the comparatively simple case of convex polytopes of arbitrary dimensions, the analogue of Euler’s theorem was discovered by Schläfli in 1850 (but not published till 1901), and independently discovered and published in the 1880’s by several mathematicians. But all these proofs—including the one in Sommerville [33]—were grossly incomplete; as noted in Grünbaum [5, p. 141], to validate these proofs one would have to show that the faces of the polytopes in question can be arranged in a special order known as *shelling*. Although it was later shown that every convex polytope can be shelled (Bruggesser-Mani [2]), the absence of this concept in the proofs shows that they were invalid in an essential aspect. (For elementary proofs of the Euler theorem for convex polytopes of all dimensions, which avoid the use of shelling, see Grünbaum [5, p. 134], McMullen-Shephard [19, p. 94].)

Returning to the situation dealing with polyhedra, a new direction opened up with the work of Hadwiger [6]. Hadwiger observed that the Euler characteristic can be defined consistently by assigning to each compact convex set C the value $\chi(C) = 1$, and proceeding to extend this to the “Konvexring” (family of sets each of which is a union of finitely many compact convex sets) by using the valuation property. Later, Klee [12] simplified and generalized this approach by putting it in a lattice setting, and observing that it also worked for unions of open convex sets (of a fixed dimension) if one starts by assigning to all open convex sets the Euler characteristic 1. Without mentioning Hadwiger or Klee, Shashkin [31] gave a detailed elementary exposition of Hadwiger’s approach to the Euler characteristic, restricted to closed and bounded polyhedral sets in the plane. However, Shashkin’s assertion (on page 82) that certain types of such sets admit a unique decomposition into “components” of a particularly “simple” type is incorrect.

In a later paper, Hadwiger [7] considered polyhedral sets, defined as those admitting relatively open convex dissections, but—due to his insistence on defining $\chi(C) = 1$ for every relatively open convex polyhedron—failed to obtain the general version of Theorem 3.

The approach followed here, to assign to a relatively open convex set of dimension d the Euler characteristic $(-1)^d$, is due to Lenz [15]. An account of the relevant writings by Lenz, Groemer and others appears in McMullen-Schneider [18]. Related developments, and in particular the relations between the Euler characteristics and valuations on appropriate families of sets are presented in Schneider [28], with extensive references to earlier literature. It should be stressed that, in all these works, the definition of the Euler characteristic by a relation such as (c) in Section 3 does not lead to any result of the nature of Theorem 4. Although involving the relatively open convex elements of dissections, these formulae apply equally to *all* such dissections of the given set (just as the topological approach applies to all simplicial complexes that represent a given set); as a consequence, they do not reflect in any way the particular facial structure of a polyhedral set. In fact, it seems that objects like the k -scaffolds have not been considered in the literature at all.

Neither McMullen and Schneider [18] nor Schneider [28] mention the work of Nef [20–25], which our note parallels to some extent. Nef’s definition of polyhedral

sets differs from ours, but is equivalent to it; his approach to the Euler characteristic is the same as ours. However, we believe our definitions lead to simpler proofs, and there is a significant difference between Nef's treatment of the faces of polyhedral sets and that given here. In view of its importance in the history of the subject, it seems appropriate to devote a few lines to a discussion of this topic.

In the case of convex polyhedra it is generally accepted that the "faces" are the intersections of the polyhedra with supporting planes. The only differences between the treatments of various authors is in deciding whether to regard the intersections themselves, or their relative interiors, as "faces", and whether the convex set itself and/or the empty set should also be included. Whatever choices are made, the faces of a convex polyhedron form a well-determined family; there is a bijection between the family of relatively open faces, and the family of closed faces. However, when more general polyhedra are considered there seems to be no agreement at all! In fact, most writers seem content to avoid the topic all together.

One of the few writers devoting some attention to this question is Hajós [8]. His "polyhedra" are closed, bounded, polyhedral sets which coincide with the closure of their interiors. To find the "faces" of such a polyhedron P , he proceeds as follows: If L is a plane that meets $\text{bd } P$ in a set with non-empty relative interior, then the closure of any connected component of the relative interior of $L \cap \text{bd } P$ is a face of P . A vertex of P is any point that belongs to some three different faces such that their planes do not pass through one line. If the intersection S of two non-coplanar faces of P contains a segment, then S contains two or more vertices of P ; segments determined by these vertices, contained in S and not containing any vertices in their relative interior, are called edges of P . Hajós appears not to realize that these definitions lead to such oddities as an edge that is in the relative interior of every face that contains it, or a vertex that is a relatively interior point of every face that contains it (see FIGURE 11(a)). Rather unsurprisingly, Hajós reports no results concerning these definitions, and it is reasonable to expect that there is little hope for establishing any connection between the Euler characteristic of P and the vertices, edges and faces defined in this way. The calculation in the caption to FIGURE 11 bears out this interpretation.

Nef [1978] defines as faces of a polyhedral set P any family of disjoint, relatively open convex sets that is a dissection of P ; hence a relation analogous to Theorem 4 holds if V, E, F, C denote the numbers of "faces" of dimension 0, 1, 2, 3, respectively. However, these "faces" are, in general, not uniquely determined, and have only a limited geometric significance due to the following fact. There exist closed polyhedral sets homeomorphic to a closed ball, for which any convex dissection must use as vertices points that cannot be regarded as vertices of P in any reasonable sense. Such polyhedral sets were first described by Lennes [14]; the simplest of these, shown in FIGURE 11(b), is due to Schönhardt [29]. For this set P our definitions yield six vertices, twelve edges and eight faces. In any of the dissections of P used by Nef there are at least seven vertices (and correspondingly larger numbers of edges and faces), hence at least one of them is devoid of geometric meaning. The same polyhedron P is also a counterexample to Lemma 2 of Szabó [34], which asserts that each polyhedron (according to a definition that includes P) has a simplicial decomposition in which all vertices of the tetrahedra involved are also vertices of the polyhedron. For interesting results concerning polyhedra that lack simplicial decompositions free of additional vertices see Rupert and Seidel [27].

The results concerning polyhedral sets presented here can be extended to more general sets. For example, one could admit as "basic constituents", besides

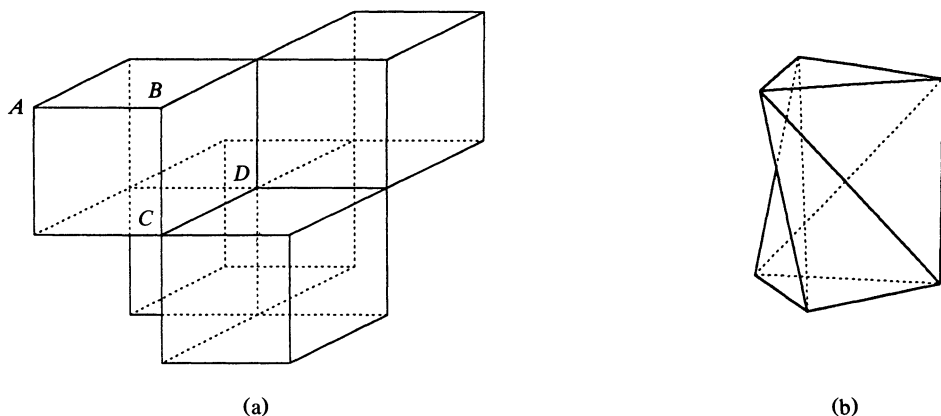


Figure 11. (a) A closed polyhedral set P , which is the union of four cubes. According to the definitions of Hajós [8, p. 37], this P has 23 vertices, 42 edges and 15 faces, hence the Euler characteristic of its boundary would be $23 - 42 + 15 = -4$. According to the definitions adopted here, $C = 4$, $V = 16$ (four vertices such as A , and 12 like B), $E = 30$ (twelve edges like AB , and six edges in the shape of two crossing open segments such as those containing the point C , each of Euler characteristic -3), and $F = 19$ (twelve faces are small squares, and one face, of Euler characteristic 7, consists of the three large squares that contain the point D), hence $\chi(\text{bd } P) = 5$, and $\chi(P) = 1$ —in agreement with the topological interpretation of the Euler characteristic. (b) A nonconvex octahedron P , which has the property that in every relatively open convex dissection of P , some vertices of the polyhedra in the dissection are not vertices of P .

relatively open convex polyhedra, also spheres, open balls, closed balls, and the (unbounded) complements of closed balls. Extending property (b) by assigning to these sets the Euler characteristic 2, -1 , 1, and -2 , respectively, and to circles and open circular disks the values 0 and 1, and considering sets obtainable as finite unions of intersections of “basic constituents”, results analogous to Theorems 1 to 4 can be obtained. Various other generalizations are also possible; their investigation is left to the reader.

We hope that the present account will lead to a better understanding of Euler’s Theorem, and possibly also to analogous results for other valuations on polyhedral sets.

REFERENCES

1. J. W. Alexander, A proof of the invariance of certain constants of Analysis Situs. *Trans. Amer. Math. Soc.* 16 (1915), 148–154.
2. M. Bruggesser and P. Mani, Shellable decompositions of cells and spheres. *Math. Scand.* 29 (1971), 197–205.
3. J. Dieudonné, Les débuts de la topologie algébrique. *Suppl. Rendiconti del Circolo Matem. Palermo*, Ser. II, No. 8 (1985), 139–153.
4. P. J. Federico, *Descartes on Polyhedra: A Study of De Solidorum Elementis*. Springer-Verlag, New York 1982.
5. B. Grünbaum, *Convex Polytopes*. Interscience, London 1967.
6. H. Hadwiger, Eulers Charakteristik und kombinatorische Geometrie. *J. reine angew. Math.* 194 (1955), 101–110.
7. ———, Erweiterter Polyedersatz und Euler-Shephardsche Additionstheoreme. *Abhandlungen Math. Seminar Hamburg* 39 (1973), 120–129.
8. G. Hajós, *Einführung in die Geometrie*. Akadémiai Kiadó, Budapest 1970.
9. J. F. C. Hessel, Nachtrag zu dem Eulerschen Lehrsatz von Polyëdern. *J. reine angew. Math.* 8 (1831), 13–20 + plate.

10. P. Hilton and J. Pedersen, Descartes, Euler, Poincaré, Pölya and polyhedra. *L'Enseignement Math.* 27 (1981), 327–343.
11. P. J. Hilton and S. Wylie, *Homology Theory, an Introduction to Algebraic Topology*. Cambridge Univ. Press 1960.
12. V. Klee, The Euler characteristic in combinatorial geometry, *Amer. Math. Monthly* 70 (1963), 119–127.
13. I. Lakatos, *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge Univ. Press 1976.
14. N. J. Lennes, Theorems on simple finite polygon and polyhedron. *Amer. J. Math.* 33 (1911), 37–62.
15. H. Lenz, Mengenalgebra und Eulersche Charakteristik. *Abh. Math. Sem. Univ. Hamburg* 34 (1970), 135–147.
16. S. A. J. L'Huilier, Mémoire sur la polyédrométrie, contenant une démonstration directe du Théorème d'Euler sur les polyèdres, et un examen des diverses exceptions auxquelles ce théorème est assujéti. *Annales de Math. pures et appl.* 3 (1812/13), 169–191.
17. J. Malkevitch, The first proof of Euler's formula. *Mitteilungen Math. Seminar Universität Giessen* 165 (1986), 77–82.
18. P. McMullen and R. Schneider, Valuations on convex bodies. *Convexity and its Applications*, P. M. Gruber and J. M. Wills, eds. Birkhäuser, Basel 1983, pp. 160–247.
19. P. McMullen and G. C. Shephard, *Convex Polytopes and the Upper Bound Conjecture*. London Mathematical Society Lecture Note Series vol. 3. Cambridge University Press 1971.
20. W. Nef, *Beiträge zur Theorie der Polyeder*. Lang, Bern 1978.
21. ———, Zur Eulerschen Charakteristik allgemeiner, insbesondere konvexer Polyeder. *Resultate der Mathematik* 3 (1980), 64–69.
22. ———, Gleichungen vom Dehn-Sommervilleschen Typ für nicht beschränkte konvexe Polytope und für Raumzerlegungen durch Hyperebenen. *Elemente der Mathematik* 35 (1980), 107–115.
23. ———, Euler's Characteristik und die Beschränktheit konvexer Polyeder. *J. reine angew. Math.* 314 (1980), 72–83.
24. ———, Zur Einführung der Eulerschen Charakteristik. *Monatshefte der Mathematik* 92 (1981), 41–46.
25. ———, Ein einfacher Beweis des Satzes von Euler-Schläfli. *Elemente der Mathematik* 39 (1984), 1–6.
26. H. Poincaré, Complément à l'Analysis Situs. *Rendiconti del Circolo Matem. Palermo* 13 (1899), 285–343.
27. J. Ruppert and R. Seidel, On the difficulty of triangulating three-dimensional nonconvex polyhedra. *Discrete and Comput. Geom.* 7 (1992), 227–253.
28. R. Schneider, Equidecomposable polyhedra. *Colloquia Math. Soc. János Bolyai*, Vol. 48 (Intuitive Geometry, Siófok, 1985), pp. 481–501 (1987).
29. E. Schönhardt, Über die Zerlegung von Dreieckspolyedern in Tetraeder. *Mathematische Annalen* 98 (1928), 309–312.
30. K. Seydel, *Geometry*. Saunders, Philadelphia 1980.
31. Y. A. Shashkin, *The Euler Characteristic*. Mir Publishers, Moscow 1989.
32. I. M. Singer and J. A. Thorpe, *Lecture Notes on Elementary Topology and Geometry*. Springer-Verlag, New York 1967.
33. D. M. Y. Sommerville, *An Introduction to the Geometry of n Dimensions*, Methuen, London 1929.
34. S. Szabó, Polyhedra without diagonals. *Periodica Mathematica Hungarica* 15 (1984), 41–49.
35. H. Weyl, Elementare Theorie der konvexen Polyeder, *Comment. Math. Helv.* 7 (1935), 290–306.

Department of Mathematics, GN-50
University of Washington
Seattle, WA 98195
grunbaum@math.washington.edu

School of Mathematics
University of East Anglia
Norwich NR4 7TJ, England
k101@cpc865.east-anglia.ac.uk

Otto Neugebauer: Reminiscences and Appreciation

Philip J. Davis

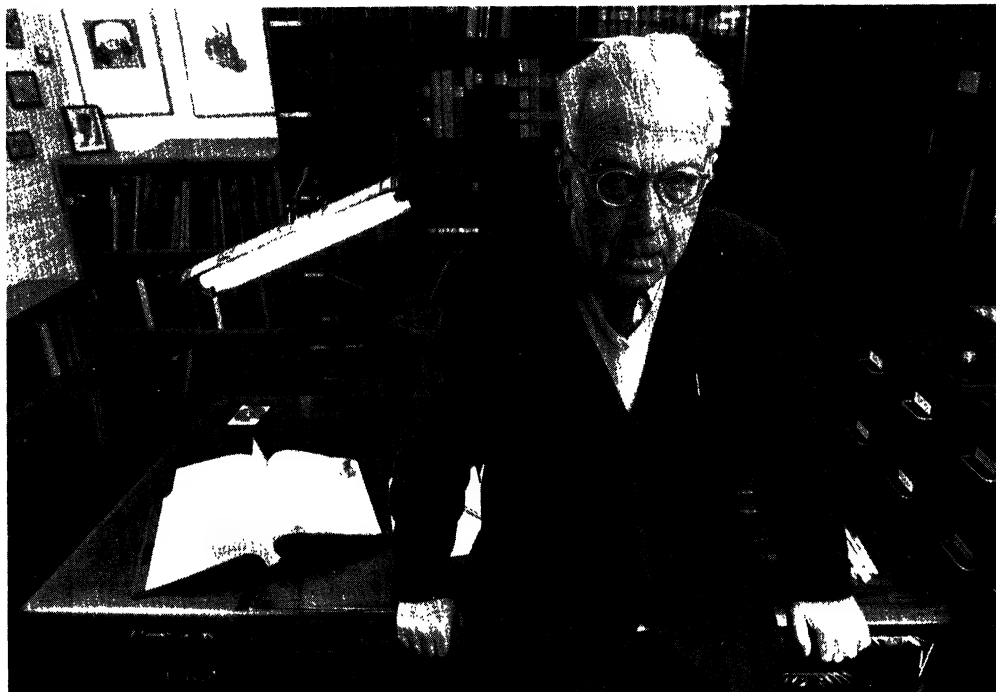
I first met Otto Neugebauer in the fall of 1963 when I joined the faculty of Brown University. For many years prior to his return to the Institute for Advanced Studies in the mid '80s, I ate lunch with him fairly regularly in the Brown cafeteria known as the Ivy Room. We were not infrequently joined by other members of the History of Mathematics Department or their visitors. The conversation was relaxed but lively, often very general, and was terminated promptly when the last person had finished lunch. Neugebauer was not one to twiddle his spoon leisurely in a second cup of coffee.

He had that soft outward appearance and low-decibel manner that I associate with Austrians. Inwardly, he held firm opinions and prejudices, and would occasionally burst out in anger and irritation. Not unlike Mark Twain, he perceived the human world as consisting largely of fools, knaves, and dupes; and when he was overwhelmed by this perception, he took refuge in his love of animals which was tender and deep.

He had his roster of The Greats in his profession, and though his ratings were not as finely tuned as those of G. H. Hardy, anyone who ate lunch with him would find out after a week which of the great names were really great and which were asses. As regards the past, he thought that Copernicus was overrated—he called him Koppernickel. Kepler was much better, and he loved Arthur Koestler's popularization of Kepler in *The Sleepwalkers*. Ptolemy was a great hero. As regards contemporaries, he expressed his views candidly.

He could not abide philosophy; he thought it a great waste of time and rarely discussed it. I would guess, though, that his largely unspoken philosophy of science was that of the logical positivism promulgated by the Vienna Circle in the 1920s. He parodied Hegel. He could not abide religious ritual or theological dogmas, and though he was a considerable student of these matters, he parodied them all mercilessly. As a relaxation, he read the *Lives of the Ethiopic Saints*, and made a dossier of the individual Devils those fellows had to contend with.

He was given to playful irony and loved Anatole France's ironic fantasies. I recall his being greatly amused by this: he said he found in some ancient Near Eastern medical text that crocodile droppings were prescribed for such and such a medical condition. The patient applied it without success and went back to the practitioner with a complaint. The practitioner asked "What was the sex of the crocodile?" "A male, I think." "Then try with a female." To this story, Neugebauer commented with a twinkle in his eye, "You see, the ancients knew all about hormones!"



Otto Neugebauer in his office at Brown. Photograph by John Forasté/Brown University

The work that made him world famous was, of course, the reconstruction of, the understanding of, the interpretation of ancient scientific texts. The text was the thing. Correspondingly, I would guess that his philosophy of history, if he ever expressed it openly, would have been that of Ranke: “*die Vergangenheit wie es eigentlich gewesen*” (the reconstruction of the past as it really was).

I suspect that his philosophy of mathematics (rarely made explicit) was that mathematics exists and has a reality independently of the people who created it. After all, for a number of years he was associated with Richard Courant and the Goettingen group. I could see at the very beginning of our friendship that his Platonic idealism was far removed from my philosophy of mathematics, and in the interests of that friendship, I stayed away from the topic at lunch time.

In line with a pervasive skepticism, and in spite of his appreciation of Koestler’s reconstruction, he was suspicious of and had little patience with attempts to link the history of mathematics with general history. He relegated such attempts pretty much to the world of the unprovable, at best, or to the world of fiction, at worst. The text, the symbol, and the recoverable platonic meaning thereof; these were his touchstones and his standards.

And yet, he knew that history was more than text, that it could not be reduced to a few symbols. He once remarked “If you never heard the sound of Nazi boots below you in the street, you cannot understand the history of the period.”

I’m sure he would have agreed with Macbeth, that human history “is a tale told by an idiot, full of sound and fury, signifying nothing.” The very separation he was able to make between the human world and the world of platonic ideas was a source of strength, a basis for his faith and his ability to pursue his work with gusto to an advanced age.

Human life and character display the particular and the universal. But as no man is an island, what seems most striking in an individual is often characteristic of a generation. Otto Neugebauer was an extraordinary individual; he also served as the emblem of an age of scholarship.

*Applied Mathematics Department
Brown University
Providence, RI 02912
am188000@brownum.brown.edu*

Correction to the Temple / Tracy article "From Newton to Einstein,"

June / July 1992.

Richard Kavinoky, Mathematics Dept., U.C. Davis

The authors' main result in Te/Tr is that the perihelion¹ shift $\Delta\theta$ in one revolution of Mercury about the Sun can be estimated by $2\pi\epsilon$, where ϵ is a small dimensionless constant determined by the Sun-Mercury system.² They use an asymptotic expansion (in ϵ) of the function for the reciprocal of the radius of the orbit, together with a general theorem (given on p. 521 of Te/Tr) to get $\Delta\theta = 2\pi\epsilon + \text{terms of order } \epsilon^2$. However, Godfrey L. Issacs pointed out that the theorem does **not** in fact apply.

In the spirit of Te/Tr we would like to formulate a new theorem that **does** apply. The function under consideration in Te/Tr is of the form $G_\epsilon(\theta) = F(\theta - \epsilon\theta) + \epsilon f(\theta)$, where $0 < \epsilon \ll 1$, F and f are 2π -periodic, i.e. G_ϵ is a **slight perturbation of a periodic function**. (The perihelion occurs initially at $\theta = 0$ and subsequently when $G_\epsilon(\theta) = G(0)$.) The Te/Tr theorem does not apply to their specific G_ϵ because $F'(0) = 0$. Thus for such G_ϵ (i.e. with $F'(0) = 0$) we want to know: when does $G_\epsilon(\theta)$ first return to its initial value $G_\epsilon(0)$, and what is the error in making a linear (in ϵ) estimate at this point?

Answer: If we add the conditions that $F, f \in C^4$, $f(0) = f'(0) = 0$ and $F''(0) \neq 0$, we may conclude that $\Delta\theta = 2\pi\epsilon + \text{terms of order } \epsilon^{3/2}$, where $\theta = 2\pi + \Delta\theta$ is the first point at which $G_\epsilon(\theta) = G_\epsilon(0)$. More precisely, $|\Delta\theta - 2\pi\epsilon| = C\epsilon^{3/2} + o(\epsilon^{3/2})$, where $C = 2\pi\sqrt{|(f''(0)/F''(0))|}$.

This is proved using Taylor's Theorem and a "bootstrap" type argument, i.e. we first get a crude estimate of the error, which enables us to then get a more precise estimate, etc., sort of pulling ourselves up by the bootstraps, as it were.

The specific Te/Tr function G_ϵ **does** satisfy these conditions and thus they are still justified in using $2\pi\epsilon$ as an estimate for the perihelion shift.

A particularly remarkable point is that although there may be more than one small $\Delta\theta$ satisfying $G_\epsilon(2\pi + \Delta\theta) = G_\epsilon(0)$, they **all** will satisfy $\Delta\theta = 2\pi\epsilon + o(\epsilon^{3/2})$.

$\epsilon \approx (3G/c^2a(1 - d^2)) \approx 7.8 \times 10^{-8}$, where G is the gravitational constant for the sun, a is the semi-major axis and d is the eccentricity of Mercury's orbit, and c is the speed of light.

[1] Issacs, Godfrey L., Unpublished letter to Editor, *American Math. Monthly*, June 20, 1992.

[2] "A Quantitative Analysis of the Precession in the Perihelion of the Orbit of Mercury (Addendum to the Temple/Tracy article "From Newton to Einstein," June/July 1992)," Richard Kavinoky, U.C. Davis Preprint.

From the Buffon Needle Problem to the Kreiss Matrix Theorem

Elias Wegert and Lloyd N. Trefethen

In this paper we present a theorem concerning the arc length on the Riemann sphere of the image of the unit circle under a rational function. But our larger purpose is to tell a story. We thought at first that the story began in 1962 with the Kreiss matrix theorem, the application that originally motivated us. However, our arc length question turns out to be more interesting than that. The story goes back to the famous “Buffon needle problem” of 1777.

SPIJKER’S LEMMA IN THE COMPLEX PLANE. Let r be a rational function of order n , that is, a quotient of two polynomials of degree at most n . Let S denote the unit circle $\{z \in \mathbb{C}: |z| = 1\}$, and let $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ denote the 1-, 2-, and ∞ -norms on S ,

$$\|f\|_1 = \int_S |f(z)| |dz|, \quad \|f\|_2^2 = \int_S |f(z)|^2 |dz|, \quad \|f\|_\infty = \sup_{z \in S} |f(z)|.$$

Then the arc length of the curve $r(S)$ in the complex plane, which we denote by $L_{\mathbb{C}}(r(S))$, can be represented compactly by the formula

$$L_{\mathbb{C}}(r(S)) := \|r'\|_1.$$

If r is multiplied by a constant α , $L_{\mathbb{C}}(r(S))$ changes by the factor $|\alpha|$. However, this scale-dependence can be eliminated by considering the ratio

$$L_{\mathbb{C}}(r(S)) / \|r\|_\infty. \tag{1}$$

In 1984, building on earlier work by Laptev, Strang, and Tadmor, LeVeque and Trefethen [9] observed that a bound on (1) could be used to derive a sharp form of the Kreiss matrix theorem (which we shall discuss at the end). They therefore posed the question, what is the maximum possible value of (1)?

It is easy to see that the value $2\pi n$ can be attained; just take $r(z)$ to be z^n or z^{-n} . If r is restricted to be a polynomial, it follows from Bernstein’s inequality that $2\pi n$ is the maximum possible. It is also easy to see that $2\pi n$ is the maximum value for rational functions in the special case $n = 1$ (the reader can supply the proof!). Based on these facts and on computer experiments, it was conjectured in [9] that $2\pi n$ is the maximum value (1) for all rational functions r and all n . However, only the bound $4\pi n$ was proved, and the task of eliminating this gap of a factor of 2 was presented as an Advanced Problem in this *Monthly* [10].

Just one response to the *Monthly* problem was received, from James C. Smith, of the University of South Alabama, who improved the bound to $2(2 + \pi)n$ [16].

Five years later, Marc Spijker of the University of Leiden finally settled the conjecture in the affirmative [17]:

Theorem 1. $L_{\mathbb{C}}(r(S))/\|r\|_{\infty} \leq 2\pi n$. (“Spijker’s lemma”)

SPIJKER’S LEMMA ON THE RIEMANN SPHERE. The simplicity of Theorem 1 is marred by the need for the normalization by $\|r\|_{\infty}$. In looking for a cleaner formulation one may ask, what is the analogous result for the Riemann sphere? Let \mathbb{S} denote the Riemann sphere $\{x \in \mathbb{R}^3: |x| = 1\}$, with the north and south poles corresponding to the points ∞ and 0 in \mathbb{C} , respectively, according to the usual stereographic projection, and the equator corresponding to the unit circle S . This identification of \mathbb{C} and \mathbb{S} is discussed in many books on complex analysis [1], and it is readily shown that a unit of arc length $|dz|$ at a position $z \in \mathbb{C}$ is expanded by the factor $2/(1 + |z|^2)$ in being projected onto \mathbb{S} . It follows that if $r(S)$ is considered as a closed curve on \mathbb{S} , with $L_{\mathbb{S}}(r(S))$ denoting its arc length on \mathbb{S} , then we have

$$L_{\mathbb{S}}(r(S)) := \|2r'/(1 + |r|^2)\|_1 \quad (2)$$

Now, the trivial scale-dependence has been eliminated from the problem. It makes sense simply to ask, what is the maximum possible value of $L_{\mathbb{S}}(r(S))$?

The new result of this paper is the following answer to this question:

Theorem 2. $L_{\mathbb{S}}(r(S)) \leq 2\pi n$. (“Spijker’s lemma on the Riemann sphere”)

The proof of Theorem 2 will emerge in the following pages. For the moment, we note first of all that like Theorem 1, Theorem 2 is obviously sharp, with equality attained for any r that maps S with winding number n onto a great circle of \mathbb{S} . (For example, $r(z) = z^n$ maps S with winding number n onto the equator, and $r(z) = i^n(z - 1)^n/(z + 1)^n$ maps S with winding number n onto the Greenwich meridian.) A more important observation is that for any r with $\|r\|_{\infty} \leq 1$, we have $L_{\mathbb{C}}(r(S)) \leq L_{\mathbb{S}}(r(S))$. This follows from (2), since $2/(1 + |r|^2) \geq 1$ when $|r| \leq 1$. Consequently, Theorem 2 implies Theorem 1 as a corollary. Thus Spijker’s lemma on the Riemann sphere is both simpler and stronger than Spijker’s lemma in the complex plane, and perhaps it should be considered the more fundamental result.

FROM THE NEEDLE PROBLEM TO POINCARÉ’S FORMULA. The reader has undoubtedly encountered the Buffon needle problem, published by the Comte de Buffon in 1777. Suppose a needle of length 1 is thrown at random on a plane ruled by parallel lines at a distance 1 apart. What is the probability that the needle will land in a position that crosses a line? Easy calculus shows that the answer is

$$\text{Probability of intersection} = \frac{2}{\pi}.$$

Buffon, incidentally, was the leading French naturalist of the eighteenth century and also a translator of Newton. He worked on his “problème de l’aiguille” long before publishing it as an appendix on “moral arithmetic” in his 44-volume treatise on natural history [3].

The needle problem became well known, especially among the French, and was generalized. Laplace, without referencing Buffon, solved the analogous problem for a square grid (*Théorie Analytique des Probabilités*, 1812). A more important generalization was to consider the slightly modified question: if the needle has

length L , possibly greater than 1, what is the *expected number* of intersections? The answer is easily seen to be

$$\text{Expected number of intersections} = \frac{2L}{\pi}. \quad (3)$$

And from here it is a small step mathematically, but a big one conceptually, to note that the same formula (3) is valid also for a *paper clip*. Various steps in this direction were taken by Cauchy, Lamé, and Barbier, among others [2]. In fact, if any rectifiable curve Γ of arc length L is thrown at random on the parallel grid, the expected number of intersections is (3). (A curve is rectifiable if its real and imaginary parts are functions of bounded variation [1].) The idea behind this result is that Γ can be thought of as a concatenation of infinitesimal straight segments, each satisfying (3) for an appropriate infinitesimal value of L . Now it may seem at first that the expected number of intersections for Γ should be more complicated than the sum of the expected numbers for the segments Γ is composed of, since after all, the segments do not fall on the grid independently. However, independence is not relevant unless one cares about the *efficiency* of (3) as a method for approximating π . It is a basic fact of statistics that the expectation of a sum of random variables is equal to the sum of the expectations, regardless of whether or not they are independent. This observation seems elementary to us now, but its application to the needle problem was evidently not obvious in the nineteenth century.

Taking the paper clip to be a circle of radius $\frac{1}{2}$ gives an easy way to remember Buffon's result and its generalization (3). For this choice of Γ , L is π and the number of intersections is exactly 2, no matter how the paper clip falls.

We now want to move from the plane to the sphere, a step taken as early as 1860 by Barbier [2]. Consider a "spherical paper clip"—that is, a curve Γ embeddable in the Riemann sphere. Suppose Γ is oriented at random on \mathbb{S} . What is the expected number of intersections with the equator? The answer is again essentially a matter of combining calculus with elementary statistics:

$$\text{Expected number of intersections on the sphere} = \frac{L}{\pi}. \quad (4)$$

Or one can skip the calculus and remember this result by thinking of the case in which Γ is itself a great circle. In this case $L = 2\pi$ and the number of intersections is again exactly 2 unless Γ happens to land exactly on the equator, an event of probability zero.

A final development completes this brief history. After Barbier, other mathematicians generalized these results further, including Poincaré, who referenced neither Buffon nor Barbier (*Calcul des Probabilités*, 1896 [12]). By this time it was clear that although the needle problem and its generalizations had conventionally been formulated as problems of probability, that interpretation could be dispensed with. Instead of orienting Γ at random on \mathbb{S} and asking for the expected number of intersections with a fixed equator, one can consider Γ to be fixed on \mathbb{S} and compute its arc length $L_{\mathbb{S}}(\Gamma)$ as an integral of the number of intersections with all great circles. To be precise, for any rectifiable curve $\Gamma \subseteq \mathbb{S}$ and any $x = (x_1, x_2, x_3) \in \mathbb{S}$, let $\nu(\Gamma, x)$ denote the number of points of intersection of Γ with the great circle on \mathbb{S} consisting of points equidistant from the antipodes $\pm x$. (When this number is infinite, the definition of $\nu(\Gamma, x)$ does not matter, for the set

of such points has measure zero.) One obtains the following elegant result:

Lemma 1. $L_{\mathbb{S}}(\Gamma) = \frac{1}{4} \int_{\mathbb{S}} \nu(\Gamma, x) dx$. (“Poincaré’s formula”)

The integral is taken with respect to area measure on \mathbb{S} .

Lemma 1 can be expressed in words as follows. To find the arc length of a curve on the Riemann sphere, integrate its numbers of intersections over all great circles, then divide by 4. Or, equivalently, since the sphere has surface area 4π , take the average number of intersections and multiply by π . This latter paraphrase of Lemma 1 makes plain its equivalence to (4).

Poincaré’s formula has far-reaching generalizations described in the book by Santaló [15], which the reader may consult for a wealth of related ideas as well as for the rigor lacking in the discussion above. It forms a centerpiece of the field known earlier as “geometric probability” but now as “integral geometry.”

PROOF OF SPIJKER’S LEMMA. Is it obvious now how to prove Theorem 2? All we need is the following lemma, whose proof we shall spell out though it might equally well have been left as an exercise. As above, $\nu(r(S), x)$ denotes the number of intersection points of the curve $r(S)$ with the great circle on the Riemann sphere \mathbb{S} defined by the points $\pm x$.

Lemma 2. *If r is a rational function of order n , then $\nu(r(S), x) \leq 2n$ for all $x \in \mathbb{S}$ with the possible exception of a single pair $x = \pm x_0, x_0 \in \mathbb{S}$.*

Proof: Since any point of \mathbb{S} can be rotated to any other by a Möbius transformation, leaving the set of rational functions of order n invariant, we are free to choose a particular value of x for convenience. Let us take x to be the north pole, so that the great circle in question is the equator, i.e., the image of S on \mathbb{S} . If $r(z) = p(z)/q(z)$ for polynomials p and q of degree $\leq n$, then for $z \in S$,

$$|r(z)|^2 = \left| \frac{p(z)}{q(z)} \right|^2 = \frac{p(z)\bar{p}(\bar{z})}{q(z)\bar{q}(\bar{z})} = \frac{p(z)p^*(z)}{q(z)q^*(z)},$$

where $p^*(z) := z^n \bar{p}(z^{-1})$ and $q^*(z) := z^n \bar{q}(z^{-1})$. The condition $|r(z)|^2 = 1$ is thus a polynomial equation in z of degree at most $2n$. Therefore $r(S)$ intersects the equator in at most $2n$ points, counted with multiplicity, unless it lies along the equator exactly. In the latter case it is obviously *only* the north and south poles for which the intersection number is infinite. ■

Since the surface area of \mathbb{S} is 4π and since $\frac{1}{4} \cdot 2n \cdot 4\pi = 2\pi n$, Theorem 2 is an immediate consequence of Lemmas 1 and 2.

Spijker’s original proof of Theorem 1, though derived independently, can be interpreted as a planar version of the same argument just given to establish Theorem 2. In particular, equation (6) of [17] is a kind of Poincaré formula for the complex plane, expressed in terms of lengths of one-dimensional projections instead of numbers of intersections. Apparently this formula was first worked out by Cauchy in 1832 and published by him in 1841 [5].

THE KREISS MATRIX THEOREM. What does all this have to do with the Kreiss matrix theorem? Let A be an $n \times n$ matrix, and let $\|\cdot\|$ denote the matrix norm induced by the vector norm $\|\cdot\|_2$. The Kreiss matrix theorem, originally published in 1962 [8], concerns the problem of characterizing matrices and families of matrices that are *power-bounded*. Let us define

$$p(A) = \sup_{k \geq 0} \|A^k\|, \quad r(A) = \sup_{|z| > 1} (|z| - 1) \|(zI - A)^{-1}\|.$$

The current, sharp form of the theorem reads as follows [17]:

Theorem 3. $r(A) \leq p(A) \leq enr(A)$. (“Kreiss matrix theorem”)

In words, a matrix A is power-bounded ($p(A) < \infty$) if and only if the norm of its *resolvent* $(zI - A)^{-1}$ increases at most inverse-linearly as z approaches the unit circle from the outside ($r(A) < \infty$). Moreover, the gap between $p(A)$ and $r(A)$ is a factor of at most e ($= 2.718\dots$) times n , so the same conclusion applies to families of matrices $\{A_\nu\}$ of fixed dimension that satisfy uniform bounds on $p(A_\nu)$ and $r(A_\nu)$.

The first inequality of Theorem 3 asserts that if $\|A^k\| \leq C$ for $k \geq 0$, then $\|(zI - A)^{-1}\| \leq C/(|z| - 1)$ for $|z| > 1$. This is easy to prove by making use of the power series $(zI - A)^{-1} = z^{-1}I + z^{-2}A + z^{-3}A^2 + \dots$. The more interesting inequality is the second one, which asserts that if $\|(zI - A)^{-1}\| \leq C/(|z| - 1)$ for $|z| > 1$, then $\|A^k\| \leq enC$ for $k \geq 0$. According to Tadmor’s remarks in [18] for the earlier developments, the history of successive improvements toward this constant en involves no fewer than nine steps, though the earlier authors in the list were certainly not concerned with optimizing the constant:

Kreiss ‘62:	$\sim [r(A)]^n$
Morton ‘64:	$\sim 6^n(n + 4)^{5n}$
Miller & Strang ‘66:	$\sim n^n$
Miller ‘67:	$\sim e^{9n^2}$
Laptev ‘75/Strang ‘78:	$32en^2/\pi$
Tadmor ‘81:	$32en/\pi$
LeVeque & Trefethen ‘84:	$2en$
Smith ‘85:	$e\left(1 + \frac{2}{\pi}\right)n$
Spijker ‘91:	en

History, thank goodness, stops here. It is shown in [9] that the constant en is best possible.

As the estimates have become sharper, the proofs have become mercifully simpler and have ceased to depend upon the explicit manipulation of eigenvalues and normal forms of matrices. We reproduce now the argument from [9] that shows how the constant en follows from Spijker’s lemma.

Proof of the second inequality of Theorem 3. According to the calculus of resolvents described for example in [7], the matrix A^k can be written as the Cauchy integral

$$A^k = \frac{1}{2\pi i} \int_G z^k (zI - A)^{-1} dz,$$

where G is any curve enclosing the eigenvalues of A , which must lie in $\{z \in \mathbb{C} :$

$|z| \leq 1\}$ if $r(A) < \infty$. Let u and v be arbitrary n -vectors with $\|u\|_2 = \|v\|_2 = 1$. Then

$$v^* A^k u = \frac{1}{2\pi i} \int_G z^k q(z) dz,$$

where v^* denotes the conjugate transpose of v and $q(z)$ is the function $v^*(zI - A)^{-1}u$, which can be shown to be a rational function of order n . Integration by parts gives

$$v^* A^k u = \frac{-1}{2\pi i(k+1)} \int_G z^{k+1} q'(z) dz.$$

Let the contour of integration be taken as $G = \{z \in \mathbb{C}: |z| = 1 + (k+1)^{-1}\}$. On this contour we have $|z^{k+1}| \leq e$ and hence

$$|v^* A^k u| \leq \frac{e}{2\pi(k+1)} \int_G |q'(z)| |dz|.$$

This integral can be interpreted as the arc length of q over the circle G . By a trivial change of variables it might as well be an arc length over the unit circle S . Theorem 1 therefore implies

$$|v^* A^k u| \leq \frac{e}{2\pi(k+1)} (2\pi n) \sup_{z \in G} |q(z)|,$$

and therefore, since the supremum of $|q(z)|$ on G is at most $(k+1)r(A)$, by the definition of $r(A)$, we have

$$|v^* A^k u| \leq enr(A).$$

Finally, we note that since $\|A^k\|$ is the supremum of $|v^* A^k u|$ over all vectors u and v with $\|u\|_2 = \|v\|_2 = 1$, this last inequality proves the theorem. ■

The Kreiss matrix theorem has been a fixture of numerical analysis since its appearance in 1962 and dissemination in the well-known book by Richtmyer and Morton [14]. It is one of the fundamental results available for establishing numerical stability of discrete processes.

CONCLUSION. From Buffon to Spijker to Kreiss, the pieces of our story fit together so neatly that it may seem there can be nothing more to say. Nevertheless, matters related to the Kreiss matrix theorem are subjects of active interest today, and in conclusion, we would like to mention a recent generalization of Theorem 3 and an open question.

The generalization concerns the problem of numerical stability of the “*method of lines*.” When time-dependent partial differential equations are solved numerically by discretization, it is common to simplify the process by constructing the space discretization and the time discretization independently. For example, the Crank-Nicolson formula for solving parabolic PDEs, of which the prototype is the heat equation $u_t = u_{xx}$, can be viewed as a second-order centered finite difference with respect to x coupled with the “trapezoid formula” with respect to t . In more realistic problems the space discretization might involve more complicated finite difference, finite element, or spectral approximations and the time discretization might be accomplished by any of the familiar methods for ODEs such as Runge-Kutta or Adams-Bashforth formulas [6].

According to the celebrated Lax Equivalence Theorem, the numerical solution computed by a consistent discretization of a well-posed linear partial differential equation will converge to the solution of the PDE as the mesh size shrinks to 0 if and only if the discretization is numerically stable [14]. (We ignore the effects of rounding errors.) But how does one test for numerical stability? It has recently been shown that for method of lines calculations, one can do it by a transplantation of the Kreiss matrix theorem from the unit disk to the subset of \mathbb{C} known as the *stability region* of the ODE formula [13]. One replaces the monomial A^k in the term $p(A)$ of Theorem 3 by the solution to a more general matrix recurrence relation, and the unit disk in the term $r(A)$ of Theorem 3 by the stability region. The condition for stability is that the norm of the resolvent of an appropriate spatial discretization matrix must increase at most inverse-linearly as z approaches the boundary of the stability region from the outside. For numerical analysts, to whom stability regions of ODE formulas are as familiar as simple groups are to algebraists, this result provides an easy means of applying the Kreiss matrix theorem to a wide range of practical problems. In particular it is applicable to the stability analysis of the high-accuracy numerical techniques known as *spectral methods* [4], where the matrices that arise are often far from normal and difficult to analyze by more elementary techniques.

The open question is, what happens to Theorem 3 if $r(A)$ is viewed as a constant rather than a variable? If $r(A) = 1$, then it can be shown that the *field of values* of A , that is, the set of Rayleigh quotients $u^*Au/\|u\|_2^2$, must lie in the closed unit disk. By a result due originally to Lax and Wendroff and subsequently sharpened by Halmos, Berger, and Pearcy [14], it follows that when $r(A) = 1$ we have $p(A) \leq 2$, or in other words, the factor *en* of Theorem 3 can be replaced by the constant 2, independently of n . Now, what if $r(A)$ is a constant greater than 1? For example, what can be said about $p(A)$ if $r(A) = 2$? It is known that $p(A)$ can no longer be bounded by a constant [11], but beyond this—for example, whether *en* can be improved to a quantity that grows only logarithmically in n —nothing is known.

ACKNOWLEDGMENTS. The problem of generalizing Spijker's lemma to the Riemann sphere was raised by the second author at a meeting in Oberwolfach in February, 1991. The proof presented here was devised by the first author, who is grateful to D. Stoyan for pointing out the connections with integral geometry. Subsequently A. I. Aptekarev of the Keldysh Institute of Applied Mathematics in Moscow has communicated to us a different and equally simple proof based on induction in n .

We are grateful also for advice from a number of others, especially Marc Spijker.

This paper was written during a visit by the second author to the Université Pierre et Marie Curie in Paris, across the street from the Jardin des Plantes where Buffon served as Director 250 years ago and his statue stands today.

REFERENCES

1. L. Ahlfors, *Complex Analysis*, McGraw-Hill, 1966.
2. E. Barbier, Note sur le problème de l'aiguille et le jeu du joint couvert, *J. Math. Pures Appl.*, 5 (1860) 273–286.
3. G. Buffon, *Essai d'arithmétique morale*, Supplément à l'Histoire Naturelle, v. 4, 1777.
4. C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral Methods in Fluid Dynamics*, Springer-Verlag, 1988.
5. A. Cauchy, Note sur divers théorèmes relatifs à la rectification des courbes, et à la quadrature des surfaces, *Comptes Rendus*, 13 (1841), p. 1060, reprinted in A. Cauchy, *Oeuvres*, Serie I, v. 6, pp. 369–375, Gauthier-Villars, Paris, 1888.
6. E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations*, Springer-Verlag, 1987 (v. 1) and 1991 (v. 2).

7. T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed., Springer-Verlag, 1976.
8. H. O. Kreiss, Über die Stabilitätsdefinition für Differenzengleichungen die partielle Differenzialgleichungen approximieren, *BIT*, 2 (1962) 153–181.
9. R. J. LeVeque and L. N. Trefethen, On the resolvent condition in the Kreiss Matrix Theorem, *BIT*, 24 (1984) 584–591.
10. R. J. LeVeque and L. N. Trefethen, Problem #6462, Advanced Problems, *Amer. Math. Monthly*, 91 (1984) 371.
11. C. A. McCarthy and J. Schwartz, On the norm of a finite Boolean algebra of projections, and applications to the theorems of Kreiss and Morton, *Comm. Pure Appl. Math.*, 18 (1965) 191–201.
12. H. Poincaré, *Calcul des Probabilités*, Gauthier-Villars, Paris, 1896.
13. S. C. Reddy and L. N. Trefethen, Stability of the method of lines, *Numer. Math.*, 62 (1992) 235–267.
14. R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial-Value Problems*, 2nd ed., Wiley-Interscience, 1967.
15. L. A. Santaló, *Integral Geometry and Geometric Probability*, Addison-Wesley, 1976.
16. J. C. Smith, An inequality for rational functions, Solutions of Advanced Problems, *Amer. Math. Monthly*, 92 (1985) 740–741.
17. M. N. Spijker, On a conjecture by LeVeque and Trefethen related to the Kreiss matrix theorem, *BIT*, 31 (1991) 551–555.
18. E. Tadmor, The equivalence of L_2 -stability, the resolvent condition, and strict H -stability, *Lin. Alg. Applics.*, 41 (1981) 151–159.

*Fachbereich Mathematik
Bergakademie Freiberg
Bernhard-v.-Cotta Str. 2
D-O-9200 Freiberg (Sachsen)
Germany
wegert@mathe.ba-freiberg.dbp.de*

*Department of Computer Science
Upson Hall
Cornell University
Ithaca, NY 14853
LNT@cs.cornell.edu*

PICTURE PUZZLE
(from the collection of Paul Halmos)



He made it possible to study compact
groups as if they were finite.
(see page 190.)

The book features leisurely and informative treatments of the three kinds of geometry considered: Euclidean (curvature $\equiv 0$), spherical (curvature $\equiv 1$), and hyperbolic (curvature $\equiv -1$). In each case, one begins with a study of the isometries of the simply connected space having the desired constant curvature and then proceeds with the study of general surfaces by means of a study of the discrete subgroups of the group of isometries. Of the three, hyperbolic geometry is perhaps the most interesting, since it is the geometry of “most” Riemann surfaces, and the connections with complex analysis are especially appealing; appropriately enough, it gets the lion’s share of attention. The sections labeled “Discussion” are especially noteworthy: they furnish brief heuristic introductions to some fascinating connections with other areas, including elliptic functions, modular forms, and Riemann surfaces. Later on, the book includes some basic material about the topological classification of surfaces and the fundamental group. When pursuing the “Riemannian covering” approach to geometry adopted here, it is quite natural to inquire what happens if one discards the requirement that the groups of isometries act without fixed points. This leads to the idea of a Riemannian *orbifold* of constant curvature, a space which locally looks geometrically like the quotient of a simply connected space of constant curvature by a finite group of isometries; this idea is introduced in the final chapter. It should also be noted that the book reads quite smoothly and is liberally furnished with helpful pictures.

This summary suggests an answer to the second question: Why is a new book needed? One answer is that the approach taken here is significantly different from that of most available texts written at a comparable level, and it furnishes comparatively easy access to some very beautiful mathematics. One of the frustrating things about teaching a one-quarter differential geometry course from one of the standard texts is that one scarcely has time to introduce the basic machinery before the term is over, and hence one never really gets very far into the *geometric* heart of the subject. For such a course, this book fills the bill admirably. Of course, there are inevitable drawbacks as well: the book neatly sidesteps the difficulties of teaching differential geometry by the rather brutal expedient of excising the “differential” to concentrate on the “geometry”, so that a student wishing to go further (say, to study traditional Riemannian geometry or general relativity) will be ill-equipped to do so; such a student would be well-advised to peruse do Carmo, O’Neill, Morgan, or even one of the many more advanced treatments of the subject to gain familiarity with the requisite machinery. Thus Stillwell’s book should be viewed as complementary to (rather than as a substitute for) the standard treatments, and in this role it is a welcome contribution. Most important, however, this book is tremendously valuable as a reminder of *why* geometry is so captivating in the first place.

Department of Mathematics
Dartmouth College
Hanover, NH 03755

Answer to Picture Puzzle
(p. 139)
Alfred Haar.

A Counterexample for Germain

William C. Waterhouse

In 1804, Sophie Germain (1776–1831) began a correspondence with Carl Friedrich Gauss (1777–1855). She began by writing under a male pseudonym [BD, Chap. 3], discussing topics in the *Disquisitiones Arithmeticae* [DA] that he had published in 1801. When she disclosed her identity to him, in 1807, he wrote a response that is famous for its comments on the obstacles women face in learning mathematics. But less attention has been paid to the actual mathematical content of the letter, where (apart from giving a brief account of some of his current work) Gauss points out a mistake in one of the results Germain had sent him and gives a counterexample. In a recent little note on the topic [MK], MacKinnon has drawn attention to the large size of the number involved (13 digits), saying “When I saw it I was filled with wonder and suspicion. Is Gauss being less than honest with Germain about the frequency of counter-examples? Surely he could never have found such a monster as a lowest counter-example?” MacKinnon found one smaller example but correctly noted that it has a special nature that might have made it inappropriate, and he concludes that he cannot see how Gauss could have discovered the example he gave. The question was echoed by Wagon [W, 319].

What I have written here is a kind of detective story in which we try to discover what really happened. We do not have the original note by Germain, so we must first reconstruct the relevant part of it from what Gauss says. Then we can go on to figure out what reasoning would have led Gauss to a counterexample. There will be a fair amount of evidence for the validity of our reconstruction; in particular, his actual counterexample will be the first one that this method would produce. I shall summarize most of the mathematics needed, so I hope that readers can also use this paper as a sort of offbeat introduction to number theory.

1. THE COUNTEREXAMPLE AS GAUSS GAVE IT. Here is the relevant part of Gauss’s letter, in my translation from the original French [G, 70–74].

The taste for abstract science in general, and especially for the mysteries of numbers, is very rare; this is no surprise, as the enchanting charms of this lofty science only reveal themselves in their full beauty to those with the courage to go deeply into it. Women, by our customs and prejudices, must encounter infinitely more obstacles and difficulties than men do to acquaint themselves with these thorny investigations; and when a person of that sex is nonetheless able to break through these barriers and penetrate the most hidden secrets, she must undoubtedly have the most noble courage, quite extraordinary talent, and superior genius. Your favoring this science, which has added so much beauty and joy to my life, reflects honor upon it; nothing could give me a more flattering and unambiguous proof that its attractions are not chimerical.

The learned notes with which your letters are so richly filled have given me countless pleasures. I have studied them with attention, and I am struck by the ease with which you have penetrated all branches of Arithmetic and perceived how to generalize and improve them. I beg you to take it as a proof of my attention if I dare to add a remark on one point in your last letter.

It seems to me that the converse proposition, that is, ‘if the sum of the n th powers of two numbers is of the form $hh + nff$, the sum of the numbers themselves will be of that same form,’ is stated a bit too generally. Here is an example where this rule fails:

$$\begin{aligned} 15^{11} + 8^{11} &= 8\,649\,755\,859\,375 + 8\,589\,934\,592 \\ &= 8\,658\,345\,793\,967 = 1\,595\,826^2 + 11 \times 745\,391^2. \end{aligned}$$

Yet $15 + 8 = 23$ cannot be reduced to the form $xx + 11yy$.

The same is true for the proposition: if one of the factors of the formula $yy + nzz$ (where n is a prime number) is of the form $(1, 0, n)$, the other necessarily belongs to the same form. Your proof only shows that no other *indeterminate* form, besides those equivalent to $(1, 0, n)$, can give the product $(1, 0, n)$ when multiplied by the form $(1, 0, n)$; but this proof does not carry over to specific numbers. For determinant $-n$, let C be any class of forms that it is not equivalent to the principal class or to any *anceps* class; let D be the class resulting from duplication of C (which will be different from the principal class); and finally let D' be the class opposite to D . It follows that the composition $C + C + D'$ yields the principal class. Thus if the two numbers f, g can be represented by a form of the class C , and the number h can be represented by a form of the class D' , the product $fg \times h$ can be reduced to $(1, 0, n)$; but it is easy to see that fg reduces not only to D or D' but also to $(1, 0, n)$. Thus here we have a case where one factor fg and the product $fg \times h$ are of the form $(1, 0, n)$ without the other factor having to be of that form. One can also see easily that the first factor must be composite; otherwise, the proposition would be true. In the example above, the factor $(15^{11} + 8^{11})/23$ contains the divisor 67.

2. RECONSTRUCTION OF THE ORIGINAL ASSERTIONS. There are some good clues in this letter that allow us to reconstruct the context of the original assertions. First, the end of the paragraph about forms refers back to the numerical example. Hence we can suppose that the mistake analyzed in that paragraph was in fact the fault in the argument for the earlier assertion. Furthermore, as Gauss explicitly cites the second result as assuming n prime, we can suppose that Germain also made that assumption in the first proposition. Indeed, as MacKinnon pointed out, there are many simple counterexamples otherwise.

The second clue is that Gauss describes the faulty result as a *converse* proposition. Thus presumably there was an earlier result stated the other way around; and since Gauss did not object to it, that result must be true. The hypotheses, therefore, must make the following statement true:

If $a + b = x^2 + ny^2$ for some x and y , then the same is true of $a^n + b^n$.

We already know we want n prime, and the values $a = 1, b = 2$ show that we want $n > 2$. It is (and was then) well known that the product of two numbers of the form $x^2 + ny^2$ is again of that form; explicitly,

$$(x^2 + ny^2)(x_1^2 + ny_1^2) = (xx_1 - nyy_1)^2 + n(xy_1 + yx_1)^2. \quad (1)$$

For odd n , we know that $a + b$ divides $a^n + b^n$ formally, and so we probably want to see when the quotient has the form $x^2 + ny^2$.

Now we can begin to get our bearings. Suppose n is an odd prime. In modern terms, Gauss had shown that the field generated over the rationals by the n -th roots of unity contains $\sqrt{\pm n}$; here we must take the plus sign when n is of the form $4k + 1$ and the minus sign when n is of the form $4k + 3$. One explicit consequence of this is in [DA, Article 357], where Gauss derives a decomposition

$$4 \cdot \frac{x^n - 1}{x - 1} = Y^2 \mp nZ^2. \quad (2)$$

These Y and Z are polynomials in x with integer coefficients, and they can be computed explicitly for any particular odd prime n . Formula (2) suggests that the direct theorem should assume that n is of the form $4k + 3$. Indeed, for (say)

$n = 5$, we can easily check that neither $3^5 + 2^5 = 275$ nor $2^5 + (-1)^5 = 31$ can be written as $Y^2 + 5Z^2$, though $3 + 2$ and $2 + (-1)$ can be. The direct theorem then must have been as follows.

Theorem A. *Let n be a prime of the form $4k + 3$. If $a + b = x^2 + ny^2$ for some x, y , then the same is true of $a^n + b^n$.*

The case $n = 3$ is special, because the quotient is already quadratic (see [MK]), so let us see how to prove this theorem for $n > 3$. The *Disquisitiones* notes that the highest term in Y is $2x^{(n-1)/2}$, and Z has no term that high. Setting $x = 0$, we have $4 = Y(0)^2 + nZ(0)^2$; hence the constant term in Y is ± 2 and the constant term in Z vanishes. Letting x approach 1, we get $4n = Y(1)^2 + nZ(1)^2$, so $Y(1) = 0$ and $Z(1) = \pm 2$. Hence an even number of terms with odd coefficients occur in each of Y and Z .

Now set $x = -a/b$; clearing denominators, we get a polynomial identity expressing $4(a^n + b^n)/(a + b)$ as a square plus n times a square. For $n = 11$, for instance, the formula in (2) is given explicitly in [DA], and it yields the identity

$$4 \cdot \frac{a^{11} + b^{11}}{a + b} = (-2a^5 + a^4b + 2a^3b^2 + 2a^2b^3 + ab^4 - 2b^5)^2 + 11(a^4b - ab^4)^2. \quad (3)$$

Our analysis of highest terms and constant terms shows in general that the terms in the squared quantities involving purely a or b are even, and there are an even number of others having odd coefficients. It follows that, whether a and b are even or odd, the two squares are even. Hence we can divide to get $(a^n + b^n)/(a + b)$ as a square plus n times a square. The theorem then follows from (1).

This proof was well within Germain's grasp. Her first letter to Gauss [S, 298–302; BD, 21] specifically singled out the decomposition of $(x^n - 1)/(x - 1)$ for praise, and at the end of her life she published a further note on that topic [G]. Thus we can be fairly sure that Theorem A was indeed her direct theorem. The converse theorem therefore must have read something like this:

(Supposed) Theorem B. *Let n be a prime of the form $4k + 3$. If $a^n + b^n$ is of the form $x^2 + ny^2$, the same is true of $a + b$.*

Germain may or may not have realized that even numbers pose a special problem. For $n = 7$, for instance, 2 is not of the form $x^2 + 7y^2$, and yet we have $2^7 + 0^7 = (4)^2 + 7(4)^2$. A related counterexample with $n = 7$ and a and b both positive was found by MacKinnon [MK, 350]. More generally, if n is of the form $7 + 16k$, we have $8(1 + 2k) = 1 + n$. We can then write

$$(2)^n + (4k)^n = 2^{n-3} \cdot 8(1 + 2k) \cdot \frac{1 + (2k)^n}{1 + 2k}.$$

By (1) and Theorem A, this quantity is of the form $x^2 + ny^2$; but $2 + 4k$ is not, as it is less than n and is not a square. This difficulty, however, turns out to be restricted to the prime 2. Indeed, as we shall see, Theorem B is true for $n = 7$ so long as $a + b$ is odd. To give the theorem the benefit of the doubt, we should therefore assume that $a + b$ is odd.

To simplify a bit more, we can note that factors equal to n are irrelevant. Indeed, as $n = (0)^2 + n(1)^2$, we know by (1) that a product nr is of the form

$x^2 + ny^2$ whenever r is of that form. Conversely, if $x^2 + ny^2 = nr$, then we see that n must divide x . Setting $z = x/n$, we get $r = y^2 + nz^2$. Hence we may as well assume that $a + b$ is not divisible by n . The same then will be true of $a^n + b^n$, as $a^n + b^n \equiv (a + b)^n \pmod{n}$. (That is, the difference of the sides is divisible by n .) Thus we may concentrate on the following version of the converse:

(Supposed) Theorem B'. *Let n be a prime of the form $4k + 3$, and suppose $a + b$ is not divisible by 2 or n . If $a^n + b^n$ is of the form $x^2 + ny^2$, the same is true of $a + b$.*

3. QUADRATIC FORMS AND THEIR COMPOSITION. To understand the discussion in Gauss's letter, we need to review a bit of his theory of quadratic forms. I shall use the notation common nowadays (see for instance [D], [B], [BS], or [C]); it is slightly different from that of Gauss, but it does not introduce any serious conceptual differences.

The *forms* are polynomials $aX^2 + bXY + cY^2$ with a, b, c all integers. An integer t is *represented* by such a form if $t = ax^2 + bxy + cy^2$ for some integers x, y . Such a form is *equivalent* to the forms obtained as

$$a(pX + qY)^2 + b(pX + qY)(rX + sY) + c(rX + sY)^2$$

where p, q, r, s are integers and $ps - qr = \pm 1$ (the condition allowing us to reverse the integral change of variables). Those forms obtained by such a change of variables with $ps - qr = 1$ are called *properly* equivalent. Since the change of variables is invertible, a number represented by one form is also represented by all equivalent forms. The number $D = b^2 - 4ac$ is called the *discriminant* of the form; it is easy to compute that equivalent forms have the same discriminant. A form is called *primitive* if there is no nontrivial common divisor of its coefficients, and again this property is preserved under equivalence.

When the form is positive for all nonzero X and Y , like $X^2 + nY^2$, then D is negative. There is then a straightforward procedure (as efficient as the Euclidean algorithm) to reduce each form to a properly equivalent form satisfying the inequalities $-a \leq b < a$ and $a \leq c$, with $a \neq c$ when $b \neq 0$. Using this, you can easily see that there are only finitely many different proper equivalence classes with a given discriminant D . In fact, it is also true that no two of these "reduced" forms are properly equivalent, and so for any D we can routinely determine the different proper equivalence classes. For instance, if $D = -44$, there are four classes, corresponding to the reduced forms $X^2 + 11Y^2$, $3X^2 + 2XY + 4Y^2$, $3X^2 - 2XY + 4Y^2$, and $2X^2 - 2XY + 6Y^2$. The first three are primitive.

The simplest way to get a change of variables with $ps - qr = -1$ is to change the sign of one variable; this amounts to changing the sign of the central coefficient b , and the result is called the *opposite* form. If two forms are properly equivalent, so are their opposites, and thus we have an operation on classes. Forms in the class of the *principal* form $X^2 + nY^2$ (which is $(1, 0, n)$ in Gauss's notation) are properly equivalent to their opposites. If forms in another class have this property, the class is called *ambiguous*, or (in Gauss's Latin) *anceps*.

In the previous section, we wrote out a formal identity (1) giving the product of $(x^2 + ny^2)$ and $(x_1^2 + ny_1^2)$. For $D = -44$, here are two more identities:

$$(3x^2 + 2xy + 4y^2)(3x_1^2 + 2x_1y_1 + 4y_1^2) = 3r^2 - 2rs + 4s^2 \quad \text{with} \\ r = xx_1 + 2xy_1 + 2yx_1 \quad \text{and} \quad s = -xx_1 + xy_1 + yx_1 + 2yy_1, \quad (4)$$

and

$$(3x^2 + 2xy + 4y^2)(3x_1^2 - 2x_1y_1 + 4y_1^2) = u^2 + 11v^2 \quad \text{with} \\ u = 3xx_1 - xy_1 + yx_1 - 4yy_1 \quad \text{and} \quad v = xy_1 + yx_1. \quad (5)$$

We can think of these as ways of composing two forms with the same D to get another one. Such expressions exist in general, but there is a very subtle difficulty involved. Clearly, for instance, we could take (4) and change the sign of y_1 , getting a valid identity with corresponding changes of signs in the expressions for r and s . Thus we would get expressions of the product in (5) by the two inequivalent forms $X^2 + 11Y^2$ and $3X^2 - 2XY + 4Y^2$. Gauss discovered that if we put suitable sign restrictions on the coefficients in the expressions of the new variables like r and s , then in fact the proper equivalence class of the form giving the product will depend only on the proper equivalence classes of the two factors. Thus we get a *composition* of proper equivalence classes. (Formulas (4) and (5) satisfy the restrictions.) A composite of primitive classes is primitive, and in modern terms the primitive proper equivalence classes form a commutative group under composition. The identity element is given by the principal class, and the opposite of a form is in the inverse class. Composing a form with itself is what Gauss called *duplication*. Formulas (4) and (5) show that the group for $D = -44$ is cyclic of order 3.

Finally, we need one fact about representations.

Lemma. *Let M be relatively prime to $4n$, and write $M = K^2M_0$ where M_0 has no repeated factors. Then M is represented by some form of discriminant $-4n$ if and only if there is some number r with $r^2 + n$ divisible by M_0 .*

Proof: If such an r exists, then the form $M_0X^2 + 2rXY + ((r^2 + n)/M_0)Y^2$ has discriminant $-4n$ and represents M (with $Y = 0$). For the converse, say $M = ax^2 + bxy + cy^2$ with $b^2 - 4ac = -4n$. Clearly then b is even. Let d be the greatest common divisor of x and y , and find s and t with $sx + ty = d$. Direct computation using our expression for M shows that

$$M(as^2 - bst + ct^2) = (s(xb/2 + yc) - t(xa + yb/2))^2 \\ + (ac - b^2/4)(sx + ty)^2.$$

Dividing by d^2 and observing that $ac - b^2/4 = n$, we see that n plus a square is divisible by M/d^2 and hence by M_0 .

The converse part of this argument was given right at the start of Gauss's treatment of forms [DA, Art. 154]. Still earlier material [DA, Art. 105] shows that such an r exists if and only if, for every prime p_i dividing M_0 , there is some r_i with $r_i^2 + n$ divisible by p_i . A simple count of powers now shows the following result:

Theorem C. *Let $M = BC$ be a number relatively prime to $4n$. If M and B are represented by forms of discriminant $-4n$, so is C .*

This may well have been in Germain's note. Observe that every number represented by a non-primitive form has a factor in common with the discriminant, and so only primitive forms will be candidates for representing M . If there is just one proper equivalence class of primitive forms (necessarily the principal class), then of course it follows that when M and B are of the form $x^2 + ny^2$, the same is true of the other factor C . This is true for $n = 3$ and for $n = 7$, and thus Germain's Theorem B' is true when n is 3 or 7.

4. LOCATING THE ERROR. The lemma above shows that M_0 will be represented by a form of discriminant $-4n$ precisely when the primes in it are so represented. When a prime number (different from 2 and n) is represented by a form of discriminant $-4n$, it is represented by forms in a unique equivalence class [DA, Art. 168]; but that usually means forms in two (inverse) proper equivalence classes. Occasionally, of course, only one proper class occurs; this happens precisely when the class is (principal or) ambiguous. We can use the representations of the primes to build up a representation of a general M by compositions. Since we can choose either of the two classes (when they are distinct), we usually get several different classes of forms that represent the same number M . It is the distinction between equivalence and proper equivalence that makes this happen, and thus it is one of the more subtle parts of the theory. And it was here that Germain made her mistake. Gauss says explicitly that she tried to prove something like this:

(Supposed) Theorem D. *If $M = BC$ is prime to $4n$ and both M and B are represented by $X^2 + nY^2$, then so is C .*

As we saw, she could correctly have proved that C is represented by some primitive form of discriminant $-4n$. To derive Theorem D (and thus Theorem B'), she would then have to show that this form was in the principal class. It appears that she tried to do this using the group structure on the classes. As Gauss says, it is indeed true that there is no *formal composition* for any other form; that is, if we have an identity

$$f(x, y)(x_1^2 + ny_1^2) = u^2 + nv^2$$

with u and v bilinear combinations of x, y, x_1, y_1 as before, then $f(x, y)$ must be in the principal class [DA, Art. 249]. But because representations of the same number can be built up to come from different classes, the formal argument fails to establish a corresponding result for specific numbers.

Following Gauss's suggestion, we can easily find explicit counterexamples to Theorem D as soon as we find an n where not all the classes are ambiguous. The first such case is $n = 11$. Take, for instance, the first three primes represented by $3X^2 + 2XY + 4Y^2$ (and hence not by $X^2 + 11Y^2$); they are

$$\begin{aligned} 3 &= 3(1)^2 + 2(1)(0) + 4(0)^2 \\ 5 &= 3(1)^2 + 2(1)(-1) + 4(-1)^2 \\ 23 &= 3(1)^2 + 2(1)(2) + 4(2)^2. \end{aligned} \tag{6}$$

In Gauss's notation at the end of Section 1, these numbers will be f , g , and h . Direct composition of the first two, as in (4), gives us

$$15 = 3 \cdot 5 = 3(-1)^2 - 2(-1)(-2) + 4(-2)^2.$$

Composition of that result with the expression for 23, as in (5), gives us

$$3 \cdot 5 \cdot 23 = (13)^2 + 11(-4)^2.$$

But we can reverse a sign to get the other expression

$$5 = 3(1)^2 - 2(1)(1) + 4(1)^2;$$

and composing this with the expression for 3, as in (5), gives

$$15 = (2)^2 + 11(1)^2.$$

Thus both 15 and $15 \cdot 23$ are represented by $X^2 + 11Y^2$, but 23 is not.

This example also shows that, if we have primes represented by $3X^2 + 2XY + 4Y^2$, we can combine them either in pairs or in triples to get numbers represented by $X^2 + 11Y^2$. By (1), then, it is easy to see that we actually have the following result:

Theorem E. *Let M be a number that is not divisible by 2 or 11. Suppose $M = p_1 \cdot p_2 \cdots p_r$ is a product of primes each representable by a form of discriminant -44 . Then M is represented by $X^2 + 11Y^2$ except when there is exactly one of the p_i not represented by $X^2 + 11Y^2$.*

5. LOCATING A COUNTEREXAMPLE. We have shown by example that the supposed Theorem D is false. It was used in the argument for Theorem B', and thus that proof is invalid, but we do not yet know that Theorem B' is false. How might Gauss have searched for a counterexample? It would be most natural to try following the same pattern; that is, we should try to take $a + b$ to be a prime represented by $3X^2 + 2XY + 4Y^2$. We know $(a^{11} + b^{11})/(a + b)$ is represented by $X^2 + 11Y^2$, and Theorem E shows that we just need to have it divisible by some other prime p that is represented by $3X^2 + 2XY + 4Y^2$. We can also recall Fermat's result (see [WL]) that such a p will have to be of the form $11k + 1$. (The point is that there is an $x \not\equiv 1$ with $ax \equiv -b \pmod{p}$; then if p divides $a^{11} + b^{11}$, we have $x^{11} \equiv 1 \pmod{p}$. But $x^{p-1} \equiv 1$ by Fermat's theorem, and so 11 divides $p - 1$.) Thus our previous work gives us the following way of searching for an example:

- 1) Take a prime represented by $3X^2 + 2XY + 4Y^2$, and write it in all ways as a sum $a + b$.
- 2) Take another prime p of the form $11k + 1$ represented by $3X^2 + 2XY + 4Y^2$.
- 3) Test whether a^{11} is congruent to $-b^{11}$ modulo p .

When the congruence holds, then both p and $a + b$ will divide $a^{11} + b^{11}$, and Theorem E will tell us that $a^{11} + b^{11}$ is represented by $X^2 + 11Y^2$.

The first primes represented by $3X^2 + 2XY + 4Y^2$ are 3, 5, and 23. The first ones also of the form $11k + 1$ are 23 and 67. Obviously $a + b = 3$ and $p = 23$ should be tried first. The only decomposition is $a = 2$, $b = 1$. It is easy to compute powers modulo a small number, and we find that $2^{11} \equiv 1 \pmod{23}$. Thus there is no example there. Similarly, we see that $4^{11} \equiv 1 \equiv 1^{11} \pmod{23}$ and $3^{11} \equiv 1 \equiv 2^{11} \pmod{23}$, so we have no examples with $p = 23$ and $a + b = 5$. We next move to $p = 67$, trying first $a + b = 3$ and then $a + b = 5$. There are still no solutions. But we have a wider range of possibilities with $a + b = 23$; and if we start from $b = 1$ and work our way up, we do find an example, at $b = 8$, $a = 15$. I repeat that checking these facts requires only computations of powers modulo 67, which are quite easy. (They are trivial if you have once computed a "table of indices" modulo 67, as described in [DA, Art. 58] and given in [DA, Table 1].) Thus relatively simple computation has led us to the following result, which we have fully established:

Theorem F. *The number $15^{11} + 8^{11}$ is represented by $X^2 + 11Y^2$, but $15 + 8 = 23$ is not.*

To confirm that we have been on the right trail, we can now observe that the one thing Gauss mentioned about his counterexample was that $(15^{11} + 8^{11})/23$ is divisible by 67.

6. COMPUTING THE COUNTEREXAMPLE. Gauss could have stopped at this point in his work, but it is not surprising that he did not. He was always fond of computation, and the counterexample will clearly be more immediately convincing if it is displayed rather than deduced. Furthermore, he had available a quite simple method for finding the expression of $15^{11} + 8^{11}$ as a square plus 11 times a square; it merely applies suitable compositions to the formula that started the whole discussion.

Let us recall what we know. First, we have

$$15^{11} + 8^{11} = 23 \cdot 67 \cdot M$$

for some integer M . (You can compute that $M = 5\,618\,653\,987$.) If we can represent M as $3X^2 + 2XY + 4Y^2$, then we can compose that representation with representations of 23 and 67 to get $15^{11} + 8^{11}$ as $X^2 + 11Y^2$.

It is in fact possible to take the value of M and solve this representation question from scratch (see Section 7.2). But we can do much better here, because we already have an expression for $67M$. Indeed, setting $a = 15$ and $b = 8$ in formula (3) above, we get

$$67M = (15^{11} + 8^{11})/23 = (-227\,723)^2 + 11(171\,780)^2. \quad (7)$$

Now we can work backwards: if we have $M = 3x^2 + 2xy + 4y^2$, and $67 = 3(3)^2 - 2(3)(4) + 4(4)^2$, composition gives us the expression $67M = u^2 + 11v^2$ with $u = 5x - 13y$ and $v = 4x + 3y$. We can solve to get $x = (3u + 13v)/67$ and $y = (-4u + 5v)/67$. Having (7), we try $u = \pm 227\,723$ and $v = \pm 171\,780$ to find values making x and y integers. Taking both u and v positive, we get the solution

$$M = 3(43\,527)^2 + 2(43\,527)(-776) + 4(-776)^2. \quad (8)$$

All we need to do now is to compose the terms differently, first combining

$$\begin{aligned} 23 &= 3(1)^2 + 2(1)(2) + 4(2)^2 \quad \text{and} \\ 67 &= 3(3)^2 + 2(3)(-4) + 4(-4)^2 \end{aligned}$$

to get

$$23 \cdot 67 = 3(7)^2 - 2(7)(-17) + 4(-17)^2.$$

Composing this then with (8), we get

$$15^{11} + 8^{11} = 23 \cdot 67 \cdot M = X^2 + 11Y^2,$$

where

$$X = 3(43\,527)(7) - (43\,527)(-17) + (-776)(7) - 4(-776)(-17) = 1\,595\,826$$

and

$$Y = (43\,527)(-17) + (-776)(7) = -745\,391.$$

Thus we have recovered exactly the example Gauss gave, and we have done it by methods all drawn from [DA].

7. SIDE ISSUES

7.1. We can find a substantially smaller example if we are willing to allow one of a and b to be negative (there is nothing gained by taking them both negative).

Indeed, if we resume the search in Section 5 with $p = 23$ and $a + b = 3$, we observe that $4^{11} \equiv 1 \pmod{23}$. Then we deduce as before that $4^{11} + (-1)^{11}$ is represented by $X^2 + 11Y^2$, while of course 3 is not. Readers might enjoy using the method from Section 6 to find the explicit representation

$$4^{11} + (-1)^{11} = 3 \cdot 23 \cdot 60\,787 = 82^2 + 11(617)^2.$$

It is also possible to find another positive example by continuing the search in Section 5 with $a + b = 23$ and $p = 67$; one finds that $a = 13, b = 10$ also works, and we have

$$13^{11} + 10^{11} = 1\,892\,160\,394\,037.$$

This is a little smaller than our previous example, but still of the same order of magnitude. The procedure of Section 6 shows us that

$$\frac{13^{11} + 10^{11}}{23} = (125\,212)^2 + 11(77\,805)^2,$$

whence

$$\frac{13^{11} + 10^{11}}{23 \cdot 67} = 3(20\,703)^2 + 2(20\,703)(-1\,669) + 4(-1\,669)^2$$

and finally

$$13^{11} + 10^{11} = (661\,539)^2 + 11(363\,634)^2.$$

7.2. We can use the value $M = 5\,618\,653\,987$ from Section 6 to illustrate the method Gauss gives [DA, Art. 322] for finding representations by $3X^2 + 2XY + 4Y^2$ from scratch. The first step is rather like that in Section 6; we observe that such an expression will give us

$$3M = Z^2 + 11Y^2 \quad \text{with} \quad Z = 3X + Y.$$

Thus it will suffice to find such representations of $3M$. There is an obvious bound

$$|Y| \leq \sqrt{3M/11} < 39\,146,$$

and any particular Y can be checked to determine whether $3M - 11Y^2$ is a square, but there are too many cases to check by hand.

Gauss's method now is a sort of sieve which he called the use of "eliminating numbers" and which is now called Gaussian exclusion [B, p. 194]. We take various small moduli E and determine conditions on Y modulo E that follow from the fact that $3M - 11Y^2$ is congruent to a square modulo E . For instance, in our case, take $E = 8$. We have $Y^2 \equiv 0, 1, 4 \pmod{8}$, while we have $3M \equiv 1 \pmod{8}$ and $-11 \equiv 5 \pmod{8}$. Thus $3M - 11Y^2 \equiv 1, 6, 5 \pmod{8}$. As only 1 among these is a square modulo 8, we see that $Y^2 \equiv 0 \pmod{8}$, which tells us that Y is divisible by 4. Similarly, if we take $E = 25$, we have $3M - 11Y^2 \equiv 6^2(1 - Y^2) \pmod{25}$, so we want $1 - Y^2$ to be congruent to a square. The squares modulo 25 are $0, \pm 1, \pm 4, \pm 6, \pm 9, \pm 11$. The condition on $1 - Y^2$ forces $Y^2 \equiv 0, 1 \pmod{25}$, so we see that either Y is divisible by 5 or $Y \equiv \pm 1 \pmod{25}$. Combining this result with divisibility by 4, we see that Y must have one of the following three forms:

$$Y = 20W, \quad Y = 100W + 24, \quad \text{or} \quad Y = 100W + 76.$$

Clearly this analysis has reduced the number of values to be checked; in the last two cases, for instance, we have $W \leq 320$. We can continue exclusion with other moduli until we feel the number of values is reasonable for checking. Gauss indicates that he was comfortable using up to 9 or 10 moduli, and that number of

steps will suffice here. For $Y = 100W + 24$, for instance, we may use moduli 3, 7, 13, 17, 19, 23, and 29; that leaves 10 values of W . The test at modulus 31 eliminates all W except 222, 239, and 304. These can be checked individually, and it turns out that none of them work. Similarly, for $Y = 100W + 76$, the moduli 3, 7, 13, 17, 19, 23, and 29 leave 12 values of W , and the test at modulus 31 eliminates all W but 7, 33, 126, 239, and 315, which we can check. It turns out that $W = 7$ works; we get

$$3M = (\pm 129\,805)^2 + 11(\pm 776)^2,$$

whence

$$M = 3(43\,527)^2 + 2(43\,527)(-776) + 4(-776)^2.$$

That, of course, is the solution we found earlier. If we want all solutions, we can attack the other case. For $Y = 20W$, we have $W \leq 1859$, and testing with the primes through 31 still leaves 25 values. The test at 37 cuts this down to 12 values, and the one at 41 then restricts W to be 538, 943, 944, 1541, or 1553. Again we can check these, and it turns out that 1553 works; we get

$$3M = (\pm 79\,019)^2 + 11(\pm 31\,060)^2,$$

whence

$$M = 3(36\,693)^2 + 2(36\,693)(-31\,060) + 4(-31\,060)^2.$$

Thus in fact there is also another representation of M . If we run through the computations for it, we get another expression for our basic counterexample:

$$15^{11} + 8^{11} = (935\,166)^2 + 11(841\,201)^2.$$

7.3. The computations in 7.2 could be shortened (using composition of forms) if we knew a factorization of M . But, as Gauss pointed out [DA, Art. 333], we can actually reverse that procedure and use our two different representations by $3X^2 + 2XY + 4Y^2$ to find factors of M . The proof of the lemma before Theorem C shows that each such representation can be rewritten to give a number whose square is congruent to -11 modulo M ; specifically, if $rX + sY = 1$, then

$$[r(X + 4Y) - s(3X + Y)]^2 \equiv -11 \pmod{M}.$$

For the solution $X = 43\,527$, $Y = -776$, the Euclidean algorithm gives $r = -153$, $s = -8\,582$, and we get the congruence

$$(1\,107\,801\,791)^2 \equiv -11 \pmod{M}.$$

The solution $X = 36\,693$, $Y = -31\,060$ similarly gives the congruence

$$(3\,256\,684\,733)^2 \equiv -11 \pmod{M}.$$

Now whenever $u^2 \equiv v^2 \pmod{M}$, each prime dividing M divides either $u + v$ or $u - v$; and so long as $u \not\equiv \pm v \pmod{M}$, neither of $u \pm v$ is divisible by M itself. Thus we can find nontrivial factors of M as the greatest common divisors of $u + v$, M and of $u - v$, M . In our case, the Euclidean algorithm shows that the greatest common divisor of $3\,256\,684\,733 - 1\,107\,801\,791$ and M is 235 159, while that of $3\,256\,684\,733 + 1\,107\,801\,791$ and M is 23 893. And indeed you can check that

$$M = 235\,159 \cdot 23\,893.$$

- [B] D. A. Buell, *Binary Quadratic Forms: Classical Theory and Modern Computations*. Springer, New York, 1989.
- [BD] L. L. Bucciarelli and N. Dworsky, *Sophie Germain: An Essay in the History of the Theory of Elasticity*. Reidel, Boston, 1980.
- [BS] Z. I. Borevich and I. R. Shafarevich, *Number Theory*. Academic Press, New York, 1966 (Pure and Applied Math., No. 20).
- [C] D. A. Cox, *Primes of the Form $x^2 + ny^2$: Fermat, Class Field Theory, and Complex Multiplication*. Wiley, New York, 1989.
- [D] H. Davenport, *The Higher Arithmetic*. Dover, New York, 1983.
- [DA] C. F. Gauss, *Disquisitiones Arithmeticae*. First edition, 1801. Standard Edition = *Werke*, vol. I. Königliche Gesellschaft der Wissenschaften, Göttingen, 1863. English Translation, Springer, New York, 1986.
- [G] C. F. Gauss, *Werke*, vol X.1. Königliche Gesellschaft der Wissenschaften and Teubner, Göttingen/Leipzig, 1917.
- [GE] S. Germain, Note de la manière dont se composent les valeurs de y et z dans l'équation $4(x^p - 1)/(x - 1) = y^2 \pm pz^2$, et celles de Y' et Z' dans l'équation $4(x^{p^2} - 1)/(x - 1) = (Y')^2 \pm p(Z')^2$. *J. Reine angew. Math.* 7 (1831), 201–204.
- [MK] N. MacKinnon, Sophie Germain, or, was Gauss a feminist?, *Math. Gazette* 74 (1990), 346–351.
- [S] H. Stupuy, *Oeuvres Philosophiques de Sophie Germain, Suivies de Pensées et de Lettres Inédites*. Paris, Paul Ritti, 1879.
- [W] S. Wagon, *Mathematica in Action*. Freeman, New York, 1991.
- [WL] A. Weil, *Number Theory: An Approach through History*. Birkhäuser, Boston, 1984.

Department of Mathematics
The Pennsylvania State University
University Park, PA 16802

Is Algebra Useful?

I would like to share with you (and if you deem useful with the AMM readership) some considerations that occurred to me upon reading an article in the latest AMM issue.

At the end of his interesting article "Two-Year Magazine Subscription Rates" (Vol 100, #1, January 1993), Underwood Dudley states that "... it is better to present mathematics to students as a glorious adventure for the mind... That it has uses is important, but incidental. Few students will use it, but all can see some of its glory."

My initial reaction to this was one of incredulity: hasn't the author noticed that the majority of people fail to identify the glory of mathematics and are in fact proud to profess a deep disdain for it? That most people tolerate the math requirements in school only because of educators' assurances that one day those mathematical skills will be useful? Expecting students to take mathematics course to appreciate the glory of the subject would not be much different from expecting us to take pottery courses to appreciate the glory of human manipulative creation. Some would enjoy it (including me) but most would object to the rationale for imposing such an experience.

My dismay, however, slowly turned to keen interest in what this statement really means: finally somebody is realizing, and admitting, that much of what we mathematicians have imposed on society as an essential part of knowledge can in fact be dispensed with; that people can live meaningful, productive and happy lives without using any mathematics beyond grade 3; that we should start thinking seriously about who needs what mathematical knowledge; and that we should start providing real rationales for each of the topics we want to have taught at any level.

Debates have raged recently between popular media and mathematicians about whether algebra is really useful. The arguments brought forward in defense of algebra were quite convincing to me, until I realized that I understood them because I am a mathematician, but they would not have convinced me if I weren't.

Hopefully some of us have initiated a move away from our ivory tower and towards the real world, where people use math, high power math, when needed, but do away with it whenever possible. Do you think that I am wrong? Then ask yourself when was the last time you used mildly advanced techniques for any purpose that had nothing to do with your job.

I thank you for your attention and I look forward to hearing from you.

Dr. Roberto Bencivenga
 Learning Assistance Centre
 Red Deer College
 Box 5005
 Red Deer, Alberta
 Canada, T4N 5H5

Cubic Equations, or Where Did the Examination Question Come From?

H. B. Griffiths and A. E. Hirst

1. INTRODUCTION. Cubic equations can give rise to a great deal of interesting elementary mathematics, and at one time 18 year-old students were expected to know something about them. This paper arises from thinking about a (very old) examination question on a cubic, and how the Examiner came to set it. One of us approached it with an interaction method, using a micro, and the other a “bare hands” traditional technique, and we then compared our efforts. This led us to find infinite families of related examination questions. As a by-product we found infinite families of explicit solutions to the Diophantine problem: *find integers p, q for which $4p^3 - 27q^2$ is the square of an integer.*

One measure of the importance of cubic equations is suggested in the following paragraph from the book [2] (p. 95) of the physicist Richard Feynman, who was commenting on some educational attitudes he had met while on a visit to Greece in 1980: the italics are ours.

They were very upset when I said the development of the greatest importance to mathematics in Europe was the discovery by Tartaglia that you can solve a cubic equation: although it is of little use in itself, the discovery must have been psychologically wonderful *because it showed that a modern man could do something no ancient Greek could do.* It therefore helped in the Renaissance, which was freeing man from the intimidation of the ancients. What the Greeks are learning in school is to be intimidated into thinking they have fallen so far below their ancestors.

Tartaglia’s method dates from 1530, more than 1000 years after the ancient Greeks. There is a lot of interesting history connected with all this, but that is not our concern here, except to recommend the general topic as a good one for students in higher education.

The stimulus for the present work arose from the need to discuss with Honours Mathematics students the point raised on p. 330 of Griffiths-Howson [3]. There, two examination questions aimed at 18 year-olds are printed, on cubic equations; the first was set (as a “Higher Certificate” question) in 1910, the second in 1972, and they are given as examples of resistance to change in the examination system—but the authors were not advocating the complete disappearance of this type of work!

One change *is* displayed in the questions: the second is more structured and depends on the explicit use of symmetric functions of the roots of the cubic. The symmetry can be used by the examinees as a check on their manipulative accuracy. But the first, older, question makes one wonder how the Examiner found the numbers in it so that it would work *at all.*

The question was: *Prove that, if α, β, γ are the three roots of the equation $x^3 - 21x + 35 = 0$ then $\alpha^2 + 2\alpha - 14$ will be equal to β or γ .*

Our interest here is the question: *How did the Examiner find the coefficients $-21, 35, 1, 2, -14$? Also, what other coefficients are possible?* The obvious approach to solving the examination question is to write X for $\alpha^2 + 2\alpha - 14$ and then show that

$$(X - \beta)(X - \gamma) = 0 \quad (1.1)$$

using a knowledge of such facts as $\alpha + \beta + \gamma = 0$.

There seems to be no way of checking possible arithmetic errors, no “conceptual” way through the question. Great manipulative ability and accuracy would be needed by the examinees; they would be unable to detect a printing error in the question.

We can, however, use the same approach in the general case, with a view to designing the original question. [Is it the most likely one a person would use in 1910?] Thus we start with the cubic polynomial

$$C(x) \equiv x^3 - px + q, \quad (1.2)$$

and then ask for which quintuples (p, q, a, b, c) we have a quadratic

$$Q(x) \equiv ax^2 + bx + c \quad (1.3)$$

such that, for any root α of C , $Q(\alpha)$ is one of the other roots of C . It turns out that, to have a sensible problem we need the quintuple (p, q, a, b, c) to be real, with

$$p \neq 0 \neq a, \quad \Delta = 4p^3 - 27q^2 > 0 \quad (1.4)$$

so that the three roots α, β, γ are all distinct and real. (Δ is called the *discriminant* of the cubic.) We then call (C, Q) a *related* pair, and show:

Theorem 1.1. *All related pairs (C, Q) can be obtained by the following process. Choose real non-zero numbers a, σ . Then take $b = (\sigma - 1)/2$, $c = (s - \sigma^2)/2$ with $s = -1$ or -3 ; and then*

$$p = \frac{(s - \sigma^2)(s - 3\sigma^2)}{4a^2(1 + \sigma^2)}, \quad q = \frac{\sigma(s - \sigma^2)(2 + s - \sigma^2)}{4a^3(1 + \sigma^2)} \quad (1.5)$$

It turns out that if $s = -3$, the correspondence $Q \rightarrow C$ is a bijection, and Q permutes the roots of C ; but if $s = -1$ and (C, Q) is a related pair, then Q does not permute the roots – and if $\sigma = \pm 1$ two Q s may correspond to the same C .

From (1.5), if a, σ lie in the field \mathbb{Q} of rational numbers, the quintuple (p, q, a, b, c) is *rational*. Moreover, if $a^2 = 1$ and σ is an odd integer, then p, q, a, b, c are all integers, so we can generate an infinity of examination questions like the one above (which itself corresponds to taking $s = -3$, $\sigma = 5$, $a = 1$). Somewhat miraculously, if (1.5) is used to evaluate Δ in (1.4), an algebraic manipulative program shows that the only real values of s , for which Δ is an algebraic perfect square, are $s = -1$ and $s = -3$. Hence if p, q are integers, so is $\sqrt{\Delta}$. Thus we are able to give infinite families of solutions to the Diophantine problem of finding integers p, q for which Δ is a squared integer. More precisely, we show:

Theorem 1.2. *Let $s = -1$. Then (1.5) gives integer values for p, q, c only when $a = \pm 1$ and $b \in \mathbb{Z}$. When $s = -3$, let \overline{W} denote the multiplicative closure of the set*

$$W = \{\text{prime numbers of the form } 6n + 1, (n \in \mathbb{Z})\}. \quad (1.6)$$

Then (1.5) gives integer values for p, q, c only when $b \in \mathbb{Z}$, and either $|a| = 1$, $|a| = 3$, or $|a| \in \overline{W} \cup 3\overline{W}$.

2. ELIMINATING THE ROOTS. With $C(x), Q(x)$ as above, we now go through the same process as the examinees, and find (as with (1.1)) when (p, q, a, b, c) are such that

$$F_\alpha = (k - \beta)(k - \gamma) = 0, \quad k = Q(\alpha). \quad (2.1)$$

Assume, for the moment, that $\alpha \neq 0$. The blanket assumption that Δ in (1.4) is strictly positive will ensure that α, β, γ are real and distinct (see, for example Birkhoff & MacLane [1] p. 432, Theorem 20).

If now we multiply out (2.1), we have

$$F_\alpha = k^2 + \alpha k - q/\alpha \quad (2.2)$$

since the sum and the product of the roots of $C(x)$ are 0 and $-q$ respectively. Since $\alpha^3 = p\alpha - q$, we can express F_α as a quadratic $F_\alpha = L\alpha^2 + M\alpha + N$, where

$$L = a^2p + 2ac + b^2 + b + 1, \quad M = ap(2b + 1) - a^2q + (2b + 1)c, \\ N = -p - (2b + 1)aq + c^2.$$

The equations $F_\beta = 0 = F_\gamma$ yield 3 linear equations in L, M, N with non-zero coefficient determinant, since we are assuming α, β, γ distinct. Therefore $L = M = N = 0$.

If, however $\alpha = 0$, then $q = 0$ and by (1.2) $\beta = -\gamma = \pm \sqrt{p} \neq 0$. So with $a = 1/\sqrt{p}$, $b = 0$, $c = -\sqrt{p}$ we have $Q(0) = -\sqrt{p}$ which is one of the two remaining roots, while $Q(\beta) = Q(\gamma) = 0$, thus ensuring that the required conditions are satisfied.

To summarise: if we assume (1.4), then “all” our Examiner has to do, in order to produce the examination question, is to find a quintuple (p, q, a, b, c) that satisfies the equations $L = M = N = 0$; for then, F_α in (2.1) will be zero (and only such quintuples will work). Of course, the Examiner would hope to find a quintuple of *integers*, but we shall deal with that point later.

If for brevity we write $\sigma = 2b + 1$ then the equations $L = M = N = 0$ become respectively:

$$(i) \quad 1 + 2ac + (\sigma^2 - 1)/4 = -a^2p, \quad (ii) \quad a\sigma p - a^2q = -c\sigma, \\ (iii) \quad p + a\sigma q = c^2 \quad (2.3)$$

The Examiner’s quintuple is $(21, 35, 1, 2, -14)$, which satisfies the equations (with $\sigma = 5$). We also observe that if $\beta = Q(\alpha)$ in (1.3), then

$$\gamma = -(\alpha + Q(\alpha)) = -a\alpha^2 - (b + 1)\alpha - c. \quad (2.4)$$

So $(p, q, -a, -(b + 1), -c)$ must also be a solution, (with σ changed to $-\sigma$). Therefore we have a check on our algebra, for we see that equations (2.3) are indeed invariant under the transformation $(p, q, a, \sigma, c) \rightarrow (p, q, a, -\sigma, c)$. The work of section 3 suggests (as is easily checked), that $(p, -q, -a, b, -c)$ is also a solution.

Before going on to find solutions of the equations (2.3), we observe that since $C(x)$ has at least one real root α , then $Q(\alpha) (\neq \alpha)$ is also a real root, so the third root must be real. A corollary is therefore that, to have a real quintuple (p, q, a, b, c) , we are *forced* to assume that Δ in (1.4) is *positive*. (If $\Delta = 0$, the roots are real, but at least two coincide.) Thus $\sqrt{\Delta}$ is also real.

If, moreover, $\sqrt{\Delta}$ is *rational*, then by a standard result, (see Birkhoff & Maclane [1] Theorem 21, p. 433), β and γ then lie in the field V obtained by adjoining α and $\sqrt{\Delta}$ to \mathbb{Q} . Hence if our examiner had started with a rational cubic $C(x)$, there always is a suitable rational quadratic $Q(x)$. But, the algebraic theory does not explicitly tell us how to calculate a , b , and c . We therefore need the following work, which establishes the algebraic theory as a payoff.

3. INVESTIGATION BY MICROCOMPUTER. So, let us consider, at first, only the integer values satisfying the given conditions. Then the most useful starting point is equation (2.3)(iii), with $\sigma = 2b + 1$. This equation tells us that $c^2 = p + (1 + 2b)aq$.

A computer program was written to run through values (< 100) of p , q , a and b , and to test for c being an integer. Those values which satisfied both this condition and equations (2.3)(i) and (ii) were then printed out to give the table below.

p	1	3	7	9	19	21	37	39	7
q	0	1	6	9	30	35	84	91	7
a	1	1	1	1	1	1	1	1	3
b	0	0	1	1	2	2	3	3	4
c	-1	-2	-5	-6	-13	-14	-25	-26	-14

There is one rogue result here where $a = 3$, but in all the other cases $a = 1$. The cases where $a = 1$ appear in pairs where, in each pair, the p -values differ by 2 and the b -values are the same.

Suppose we now look at the cases where $a = 1$ and consider the first p of each pair. The method of differences suggests the following results.

$$\begin{array}{ccccc} p & q & a & b & c \\ 1 + 3n + 3n^2 & n + 3n^2 + 2n^3 & 1 & n & -(1 + 2n + 2n^2) \end{array} \tag{3.1}$$

$$\begin{array}{ccccc} 3 + 3n + 3n^2 & 1 + 3n + 3n^2 + 2n^3 & 1 & n & -(2 + 2n + 2n^2) \end{array} \tag{3.2}$$

These values satisfy equations (2.3), even if we replace n by any real number b . To summarise we have the following lemma, using the notation of Section 1. It concerns a cubic $C(x)$ and a quadratic $Q(x)$ of the forms (1.2), and (1.3).

Lemma 3.1. *Suppose that in (1.3), the quadratic $Q(x)$ has $a = 1$, and b is any specified real number. Then there are two possible values of c , for each of which we can find a cubic $C(x)$, such that (C, Q) is a related pair.* ■

When $a \neq 1$, a computer search was made, similar to that described at the beginning of the section, with $a = 1/2, 1/3, 1/4, 1/5, 1/6$. It suggested that if a and b are any real numbers with $a \neq 0$,

$$\begin{aligned} p &= 3(1 + b + b^2)/a^2, & q &= (1 + 3b + 3b^2 + 2b^3)/a^3, \\ c &= -2(1 + b + b^2)/a, \end{aligned}$$

will satisfy the equations (2.3).

These solutions correspond to those given in (3.2), and we shall refer to them as Type 2 solutions. Their form suggests that we might introduce similar denominators in the sequence of solutions given in (3.1). We call these Type 1 solutions. In

both cases it is easily verified that these *are* solutions to the equations (2.3), and we can summarise these findings in the following table.

	Type 1	Type 2	
p	$(1 + 3b + 3b^2)/a^2$	$3(1 + b + b^2)/a^2$)
q	$(b + 3b^2 + 2b^3)/a^3$	$(1 + 3b + 3b^2 + 2b^3)/a^3$)
c	$-(1 + 2b + 2b^2)/a$	$-2(1 + b + b^2)/a$)

(3.3)

Hence if (p, q, a, b, c) is a solution, then so is $(p, -q, -a, b, -c)$. Using the remark following (2.4) we may summarise:

Lemma 3.2. *For any real number b , and any real non-zero a , there are two quintuples (p, q, a, b, c) satisfying the equations (2.3). Further solutions are then $(p, q, -a, -(b + 1), c)$ and $(p, -q, -q, b, -c)$. Given b, a , the Examiner has several related pairs (C, Q) available. ■*

As we shall see, the two types of solutions in (3.3) are *all* solutions of equations (2.3).

Next, by investigating Type 1 solutions when $a = 1$, it becomes apparent that the equation $C(x) = 0$ has integer roots, namely $b, b + 1$ and $-(2b + 1)$.

Observation 3.3. *In (1.3), Q need not permute the roots.*

In the case under discussion $Q(b + 1) = b = Q(-(2b + 1))$ showing that Q does not permute the roots. So, for any related pair (C, Q) , $Q(\alpha) = Q(\beta)$ iff b/a is the third root of $C(x)$; and then it is easily verified that $C(b/a)$ is always 0 or $1/a^3$ according as the solutions are of Type 1 or Type 2. Thus with Type 2, Q permutes the roots of C , but not with Type 1. An interesting sidelight in the former case: Q has an orbit of period 3, so it is an explicitly constructed quadratic which, by Li and Yorke [6], has iterations which exhibit chaos.

4. SOLVING EQUATIONS (2.3). The previous investigation starts from a and b , then finds p, q and c . But suppose we had found any real solution (p, q, a, b, c) of equations (2.3). Then, by (ii) and (iii), p and q would *have* to satisfy

$$p = \frac{c(ac - \sigma^2)}{a(1 + \sigma^2)}, \quad q = \frac{\sigma c(1 + ac)}{a^2(1 + \sigma^2)}, \quad (4.1)$$

and then substitution in (2.3)(i) requires (after simplification)

$$(2ac + t)^2 = 1, \quad t = \sigma^2 + 2. \quad (4.2)$$

Conversely, we can compute p, q from (4.1) for any choice of (a, c, σ) with $a \neq 0$; and (2.3)(ii), (iii) will be satisfied. If also (a, c, σ) satisfies (4.2) then we have obtained a quintuple (p, q, a, σ, c) satisfying all three equations (2.3).

Observe that (4.2) tells us that the only relevant quadratics $Q(x)$ in (1.3) are those for which (a, b, c) lies on one or other of the pair S^+, S^- of quadric surfaces in \mathbb{R}^3 with equations $2ac + (2b + 1)^2 = s$, where $s = -1$ for S^+ and $s = -3$ for S^- . For our Examiner, $s = -3$, but we should consider also the case $s = -1$.

Starting with s, a, σ in (4.2), c has to be $(s - \sigma^2)/2a$, and then by substitution in (4.1) we would have the equations (1.5) above. Now, for each real s , the equations (1.5) define a mapping

$$f_s: (a, \sigma) \rightarrow (p, q)$$

of the (a, σ) plane A to the (p, q) plane. Our previous investigation therefore required us to work out the *image* of f_s when $s = -3$ and -1 .

First, what could the boundary of this image be? The family of all cubic polynomials of the form (1.2) separates most basically into those with 3 real roots, and the rest. We model the former set by the subset D of the (p, q) -plane which is defined using the discriminant in (1.4):

$$D = \{(p, q) \in \mathbb{R}^2: \Delta = 4p^3 - 27q^2 > 0\} \quad (4.3)$$

with boundary the semi-cubical parabola

$$K: q = \pm 2(p/3)^{3/2}. \quad (4.4)$$

Here, D lies in the right hand half ($p > 0$) of the (p, q) -plane, as shown in FIGURE 1.

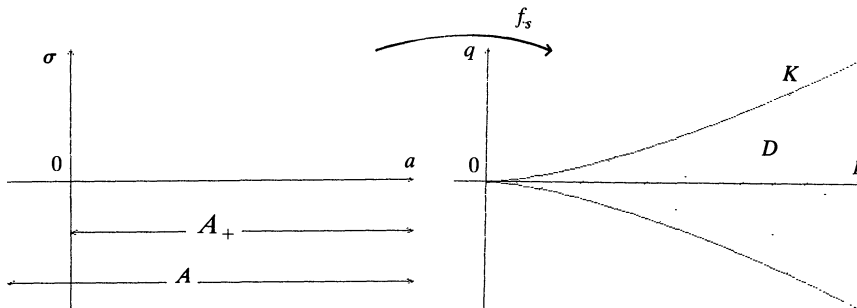


Figure 1

Now in (1.5), $p > 0$ since s and $-\sigma^2$ are negative, so we are led to guess and then prove the following:

Proposition 4.1. *If $s < 0$ and $(p, q) \in D$, (p, q) lies in the image of f_s .*

Proof: Given $(p, q) \in D$, then we must find a point (a, σ) with $f_s(a, \sigma) = (p, q)$. If $q = 0$, then $(s/2\sqrt{p}, 0)$ will do. If $q \neq 0$, let $q/p = m \neq 0$; then by (1.5), (a, σ) satisfies

$$m = \sigma(2 + s - \tau)/a(s - 3\tau) \quad (\tau = \sigma^2). \quad (4.5)$$

We solve this for a and substitute in the first equation of (1.5) to find some $\tau > 0$ satisfying

$$4p\tau(1 + \tau)(2 + s - \tau)^2 = m^2(s - \tau)(s - 3\tau)^3, \quad (4.6)$$

with m given. The difference between the left and right sides of (4.6) is $-m^2s^4 < 0$, (since $ms \neq 0$) at $\tau = 0$, and approximately $(4p - 27m^2)\tau^4 = \tau^4\Delta/p^2$ when τ is large. Therefore by the Intermediate Value Theorem, the required $\tau_0 > 0$ exists. ■

We now confine our investigation of (4.6) to the relevant cases $s = -1$ and the Examiner's $s = -3$.

Some facts about f_{-3}, f_{-1} are these, and are left as an exercise for the reader. f_{-3} is an injection (one-to-one) if we restrict its domain to the region A_+ : $a > 0$ of

A. For each $a > 0$, f_{-3} maps the line L_{a_0} : $a = a_0$ to the curve Γ_{a_0} given with parameter σ by

$$p = 3(3 + \sigma^2)/4a_0^2, \quad q = -\sigma(3 + \sigma^2)/4a_0^3,$$

so

$$dq/dp = \dot{q}/\dot{p} = -(1 + \sigma^2)/6a\sigma$$

except when $\sigma = 0$, when the tangent is vertical. The curves Γ_{a_0} are as shown in FIGURE 2.

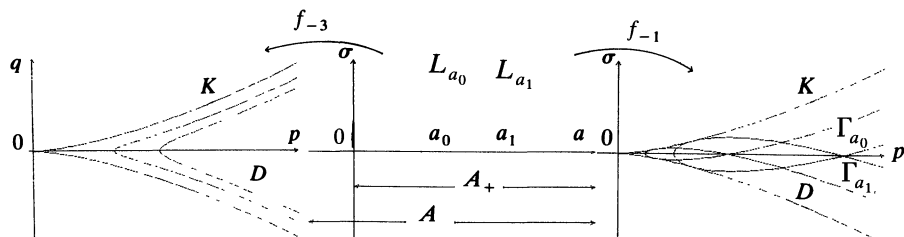


Figure 2

Thus f_{-3} : $A_+ \rightarrow D$ is a bijection, and a *diffeomorphism* if $\sigma > 0$. Following (3.3) we noted that if (p, q, a, b, c) is a solution there, then so is $(p, -q, -a, b, -c)$; hence f_{-3} is also an injection at points of A with $a < 0$. Thus, in Theorem 1.1, the correspondence $Q \rightarrow C$ is a bijection (when $s = -3$).

If again we confine the mapping f_{-1} to the half-plane A_+ of the (a, σ) -plane, we see that f_{-1} : $A_+ \rightarrow D$ is onto, but it is no longer an injection; for f_{-1} maps the line L_{a_0} to the curve Γ_{a_0}

$$p = (1 + 3\sigma^2)/4a_0^2, \quad q = \sigma(1 - \sigma^2)/4a_0^3, \quad dp/dq = (1 - 3\sigma^2)/6\sigma.$$

Thus f_{-1} maps as shown in FIGURE 2. Therefore in Theorem 1.1, if $s = -1$, then at most *two* quadratics Q can be paired with the same cubic C .

5. ALGEBRAIC FORMULAE: THE CASE $s = -3$. Here (4.6) reduces through cancellation to a *linear* equation in τ , yielding $\tau = 81q^2/\Delta$. The formulae (1.5) reduce to

$$p = 3(3 + \tau)/4a^2, \quad q = \sigma(3 + \tau)/4a^3, \quad \text{with } \Delta = 81(3 + \tau)^2/16a^6. \quad (5.1)$$

Thus we choose a, σ arbitrarily, then take $c = -(3 + \tau)/2a$ to compute p, q from (5.1) and obtain the required quintuple (p, q, a, b, c) with $b = (\sigma - 1)/2$. It is easily verified that these are exactly the “solutions of Type 2” in (3.3). Choosing the sign of the root relative to that of a , since $3 + \tau > 0$, we then have

$$a = 3p/\sqrt{\Delta}, \quad \sigma = 9q/\sqrt{\Delta}, \quad c = -2p^2/\sqrt{\Delta} \quad (5.2)$$

which confirms the statement made in Section 2 that if $p, q, \sqrt{\Delta}$ are all rational, so are a, b and c .

Integer solutions when $s = -3$. Now, our Examiner found *integers* as coefficients, and since his $p = 21$, $q = 35$, then $\sqrt{\Delta} = 63$, another integer. Certainly, if we require $p, q, a, c \in \mathbb{Z}$, then (5.1) also requires $\sqrt{\Delta}$ to be rational. The problem then arises of finding solutions of this type. In particular, for which integral p, q is $\sqrt{\Delta}$ in \mathbb{Z} ? We therefore have now a *Diophantine problem*.

Recall that $\tau = \sigma^2$ and $\sigma = 2b + 1$. If $B = b^2 + b + 1$ then from (5.1), (5.2) we obtain:

$$p = 3B/a^2, \quad q = (2b + 1)B/a^3, \quad c = -2B/a, \quad \sqrt{\Delta} = 9B/a^3. \quad (5.3)$$

Therefore, if $a = 1$, we have an infinite sequence of solutions in \mathbb{Z} as b runs through \mathbb{Z} :

$$(p, q, a, b, c) = (3B, (2b + 1)B, 1, b, -2B), \quad \sqrt{\Delta} = 9B \quad (5.4)$$

and, when $b = 2$, this gives the Examiner's solution. In Section 7 other solutions are discussed.

6. THE CASE $s = -1$. It is technically more difficult to have $s = -1$ because equation (4.6) becomes a cubic. The parametric formulae (1.5) now become, when $\tau = (2b + 1)^2$ as before,

$$\begin{aligned} p &= \frac{1 + 3b + 3b^2}{a^2}, & q &= \frac{(b + b^2)(2b + 1)}{a^3}, \\ c &= -\frac{(1 + \tau^2)}{2a} = -\frac{(1 + 2b + 2b^2)}{a}. \end{aligned} \quad (6.1)$$

It is easily checked that these are the "solutions of Type 1" in (3.3). For $\sqrt{\Delta}$ we obtain

$$\sqrt{\Delta} = (9\tau - 1)/4a^3 = (9b^2 + 9b + 2)/a^3. \quad (6.2)$$

Therefore, if $a = \pm 1$ and b runs through \mathbb{Z} , we obtain a further family of solutions to the Diophantine problem of finding (p, q) in \mathbb{Z} for which Δ is a square. In Lemma 7.1 below we show that there are no Type 1 integral solutions if $|a| > 1$.

How accidental is it that Δ is a perfect square when $s = -1, -3$? Adrian Oldknow gave us an explanation which we summarise as follows. Using an algebraic manipulation package, he substituted p, q from (1.5) in $\Delta = 4p^3 - 27q^2$ to express Δ in the form $X^2 \cdot (u(s)(1 + x)^{-1} + v(x, s))$, where $v(x, s) = 81x^2 + 9x \cdot g(s) + h(s)$, and $u(s), g(s), h(s)$ are polynomials in s of degree 4, 2, 3 respectively. Because of the denominator $1 + x$, Δ cannot be a perfect square unless $u(s) = 0$, which occurs for real s only when $s = -1, -3$. In those cases it turns out that

$$v(x, -1) = (9x - 1)^2, \quad v(x, -3) = 81(x + 1)^2,$$

and Δ is then a perfect algebraic square.

7. INTEGER SOLUTIONS AND NUMBER THEORY. The approach using a micro yielded the patterns arising out of this problem, and so we consider Types 1 and 2 (defined in (3.3)) separately, and look for values of a which give integer solutions. It is clear from (5.3) that if $B = b^2 + b + 1$, such solutions will exist whenever $a^3 | B$. The following sequence of lemmas leads to a proof of Theorem 1.2 above.

Lemma 7.1. *For Type 1 equations there are integer solutions iff $|a| = 1$.*

Proof: If p and c are integers, a divides the numerators of both p and c in (6.1) and (6.2). But if the integer a divides both $1 + 3(b + b^2)$ and $1 + 2(b + b^2)$, then $a = \pm 1$. ■

Moving to Type 2 integer solutions, we quote Corollary 25.1 of Grosswald, Chapter 7 [4]. Here, the *discriminant* of a quadratic polynomial $f(x) \equiv ax^2 + bx + c$ is $b^2 - 4ac$, and we shall take $f(x)$ to be $B(x) \equiv x^2 + x + 1$ with discriminant -3 .

Lemma 7.2. *If D is the discriminant of the polynomial $f(x)$, and if P is a prime such that $P \nmid D$, then all congruences*

$$f(x) \equiv 0 \pmod{P^r}, \quad (r \in \mathbb{N})$$

have the same number of solutions, which is in particular the number of solutions of

$$f(x) \equiv 0 \pmod{P}, \quad \blacksquare$$

The prime 3 has an anomalous role in the sequel precisely because $3|D$ in our case.

Now recall the set W in (1.6) and its multiplicative closure \bar{W} .

Lemma 7.3. *For any $P \in W$, and $r \in \mathbb{N}$, there exist Type 2 integer solutions for any a of the form P^r .*

Proof: Since $B = B(b)$ in (5.3) is odd, it cannot have 2 as a factor. Let $P \in W$. Then, working mod P , we have, since P is odd

$$1 + b + b^2 \equiv 0 \Leftrightarrow 4 + 4b + 4b^2 \equiv 0 \Leftrightarrow (1 + 2b)^2 \equiv -3$$

This means that

$$P|B \Leftrightarrow -3 \text{ is a square } \pmod{P}, \quad (7.1)$$

which happens whenever P is of the form $6n + 1$ (by Hardy and Wright [5], Th. 96), that is whenever $P \in W$. Thus $\forall P \in W, \exists b \in \mathbb{Z}$ such that $P|B(b)$. But the quadratic form $B(b)$ has discriminant -3 , to which P is prime. Therefore, by Lemma 7.2, $\forall r \in \mathbb{N}, \exists b' \in \mathbb{Z}$ such that $P^{3r}|B(b')$. Hence (5.3) gives integer solutions when $(a, b) = (P^r, b')$. \blacksquare

We leave the short proofs of the next two lemmas as an exercise for the reader.

Lemma 7.4. *Let $B = B(b)$ and let y be an odd integer prime to 3. If $y^2|B$ and $y^3|(2b + 1)B$ then $y^3|B$.* \blacksquare

Lemma 7.5. $3|B(b) \Leftrightarrow b = 3k + 1$ for some $k \in \mathbb{Z}$, in which case $9 \nmid B(b)$. \blacksquare

Lemma 7.6. *If we have a known solution of either type for a given a , then we can find an infinite sequence of integer solutions for this a .*

Proof: Already in (5.3) we have an infinite family of solutions for $a = \pm 1$. Hence the Type 1 case is clear from Lemma 7.1. For Type 2 solutions, recalling (5.3) we wish to know for which other values of a and b we have c, p and q in \mathbb{Z} . This will be true if $a^3|B(b)$, so we first look at the divisors of $B(b)$. If $|a| = 3$ we first note that there is an integer solution with $b = 4$. For then

$$B(4) = 1 + 4 + 4^2 = 21 \quad \text{and} \quad 1 + 2b = 1 + 2 \cdot 4 = 9, \quad (7.2)$$

so 27 divides the numerator of q and thus c, p and q are integers in (6.1). To get further solutions we note that $B(4 + 9x) = 21 + 81(x + x^2)$ is divisible by 3, while

$1 + 2b = 9(1 + 2x)$ is divisible by 9, so the numerator of q is divisible by 27. Thus if $|a| = 3$ there is a solution for $b = 4 + 9x$ for any $x \in \mathbb{Z}$, hence giving an infinite sequence of integer solutions.

If $|a| > 3$, suppose we have found some integer t , such that (a, t) gives a solution, (i.e., $a^3 | B(t)$). Then by a method similar to the above case with $(a, t) = (3, 4 + 9x)$, it is easy to check that $(a^3, (t + a^3x))$ is also a solution for any $x \in \mathbb{Z}$. ■

Lemma 7.7. *If $w \in \overline{W}$, then there are Type 2 integer solutions for $a = w$, and $a = 3w$.*

Proof: We prove this inductively. Let $w = p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$, where the p_i , ($i = 1, \dots, k$) are distinct primes. Then from the definition of \overline{W} , $p_i \in W$ for $i = 1, \dots, k$.

Let $x = p_1^{n_1} \dots p_{k-1}^{n_{k-1}}$, $y = p_k^{n_k}$. Then since the p_i are distinct, $\text{hcf}(x, y) = 1$, and hence $\text{hcf}(x^3, y^3) = 1$. Suppose there are solutions for $a = x$ and $a = y$. Then $\exists b_1, b_2 \in \mathbb{Z}$ such that $x^3 | B(b_1)$ and $y^3 | B(b_2)$. Then by the method of Lemma 7.6, it follows that

$$\forall u, v \in \mathbb{Z}, x^3 | B(b_1 + ux^3), \text{ and } y^3 | B(b_2 + vy^3).$$

So if we can find u, v such that

$$b_1 + ux^3 = b_2 + vy^3 = b^*, \quad \text{say,} \quad (7.3)$$

then, since $\text{hcf}(x^3, y^3) = 1$, $x^3 y^3 | B(b^*)$, thus giving an integer solution for $a = xy$. Now there exist integers m and n such that $mx^3 + ny^3 = 1$ so by choosing $u = (b_2 - b_1)m$ and $v = (b_1 - b_2)n$ it can be easily verified that equation (7.3) is satisfied, thus ensuring that there is an integer solution of Type 2 for $a = xy$.

We know by Lemma 7.3 that when $k = 1$, w yields integer solutions so, by induction, there is a Type 2 integer solution for any $w \in \overline{W}$.

Now we consider the case when $a = 3w$, $w \in \overline{W}$. We know from above that there exists a b such that $\forall h \in \mathbb{Z}$, $(w, b + hw^3)$ is a solution; and that $\forall t \in \mathbb{Z}$, $(3, 4 + 9t)$ is a solution, so we need to find integers u, v such that $4 + 9u = b + uvw^3$.

Since $\text{hcf}(9, w) = 1$, the same method used for x, y above will show that such u, v do indeed exist. Hence, for any $w \in \overline{W}$ we can find an integer solution of Type 2 when $a = 3w$. ■

We are now ready for the final step which proves Theorem 1.2.

Theorem 7.8. *All Type 2 solutions (p, q, a, b, c) , have either $|a| \in \overline{W}$, or $|a|/3 \in \overline{W}$, or $|a| = 1$ or $|a| = 3$. Moreover, if a is any of these forms, there is a Type 2 solution.*

Proof: We prove only the first sentence, as the second is essentially Lemma 7.5. Thus we shall deduce the form of a , which gives integer solutions for:

$$p = \frac{3B}{a^2}, \quad q = \frac{B(2b+1)}{a^3}, \quad c = \frac{-2B}{a}, \quad \text{where } B = b^2 + b + 1. \quad (7.4)$$

For brevity we say “by p ” instead of “by the formula for p ”. Then by p , a is odd because B is always odd. We may clearly assume $|a| > 3$.

First suppose $3 \nmid a$. Then, since a is odd, $a | B$ by c ; also by p , $a^2 | B$. Hence by Lemma 7.4 we may assert: $a^3 | B$. Further, $a^3 | 4B = (2b+1)^2 + 3$, so -3 is a square (mod a^3), and hence, by Lemma 7.2, (mod P^k) for every prime power P^k

dividing a . Therefore by Hardy and Wright [5] Theorem 96, $P \in \mathcal{W}$. Hence since $|a| > 1$, $|a| \in \overline{\mathcal{W}}$ as required.

Next suppose $3|a$, so $a = 3\alpha$ say, with α odd and > 1 since $|a|$ is odd and > 3 . Then $3 \nmid \alpha$, since by p this would imply that $9|B$ which we know from Lemma 7.5 to be false. So, repeating the above argument with α replacing a , we have $\alpha^3|B$ and hence $|\alpha| = |a|/3 \in \overline{\mathcal{W}}$, as required. ■

Finally, we describe a method for finding a suitable b in (7.3), given a in $\overline{\mathcal{W}} \cup 3\overline{\mathcal{W}}$. This depends upon the crucial observation (7.1). Choose a , and let x and y be integers such that

$$x = (a + 1)/2, \quad y^2 \equiv -3 \pmod{a};$$

then if $t = x(-1 \pm y)$, a direct computation gives

$$B(t) \equiv a \left[-\frac{(a + 3)}{2} \pm y \frac{(a + 1)}{2} \right] \pmod{a} \equiv 0 \pmod{a},$$

since both $(a + 1)$ and $(a + 3)$ are even, so the interior of the square bracket is an integer.

Having thus found a t such that $a|B(t)$, we now need a b such that $a^3|B(b)$. It can be shown that for some $n < a^2$, $b = t + na$ will do. So, for each a , the computer has only a finite number of integers to run through to find one which satisfies the required condition. This method also works with composite values of a , provided the factors of a are all in \mathcal{W} . A computer search using this method yielded the following solutions:

a	3	7	13	19	21	31	...	91
b	4	18	1036	2819	1390	6287		69267

However, it can be seen that the numbers b soon become very large. Now, $B(b)/a^3 - [B(b)/a^3]$ may be very small in comparison with $B(b)$, so an ordinary microcomputer will soon give unreliable answers. Thus, although a microcomputer is an extremely useful tool for generating patterns and prompting ideas, the theory to back up these ideas is still as necessary as ever.

REFERENCES

1. G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, Macmillan.
2. R. Feynman, *What Do You Care What Other People Think?*, Unwin, 1988.
3. H. B. Griffiths and A. G. Howson, *Mathematics: Society and Curricula*, CUP, 1974.
4. E. Grosswald, *Topics from the Theory of Numbers*, Birkhäuser, 1984.
5. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford, Fifth Edition, 1979.
6. T. I. Li and J. A. Yorke, Period three implies chaos, *Am. Math. Monthly* (82), 1975, 985–992.

Faculty of Mathematical Studies
University of Southampton
Southampton, SO9 5NH
United Kingdom
ae@uk.ac.soton.maths

NOTES

Edited by: John Duncan

On the Identity of Polyhedra

Hellmuth Stachel

Dedicated to Prof. O. Giering, Munich, on the occasion of his 60th birthday

Let S and S' be two polyhedral solids in the Euclidean 3-space E^3 . In this note it is proved that S and S' need not coincide though they share all vertices and all planes through their faces.

Let $V = \{v_1, \dots, v_m\}$ and $V' = \{v'_1, \dots, v'_m\}$ denote the sets of vertices of S and S' , respectively. $P = \{p_1, \dots, p_n\}$ and $P' = \{p'_1, \dots, p'_n\}$ are supposed to be the sets of planes containing the faces of S and S' . FIGURE 1 shows two different polyhedra S and S' with equal vertex sets $V = V'$. This is caused by different triangulations of the skew quadrangle $v_2v_3v_6v_5$. With the example in FIGURE 2 we obtain that also $P = P'$ does not imply $S = S'$. Here S (left) is the intersection and S' (right) the union of two congruent "houses". The following example (FIGURE 3) reveals that the two polyhedra can even be different though both the vertex sets and the plane sets are equal:

We start with a regular dodecahedron D . There are five inscribed cubes C_1, \dots, C_5 . And each cube C_i contains a left tetrahedron L_i and a right one R_i . The edges of the left tetrahedra L_1, \dots, L_5 of D can be defined in the following way: Start at any vertex v_0 of D along any edge e_1 . At the endpoint v_1 take the left¹ edge e_2 and at the next vertex v_2 take the right edge e_3 . Then v_0 and the endpoint of v_3 of e_3 determine an edge of any left tetrahedron. Commutation of left and right gives the right tetrahedra edges. Let S denote the union of all left tetrahedra of D (FIGURE 3, left) and S' be the union of all right tetrahedra

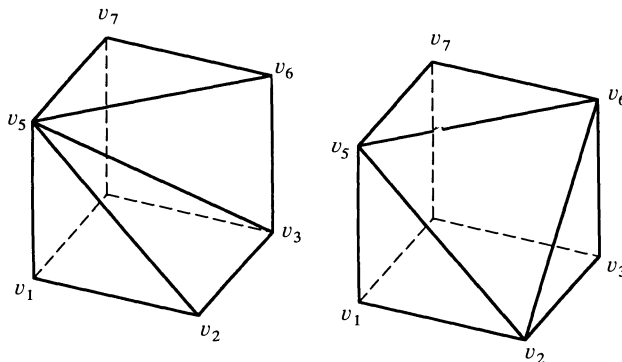


Figure 1.

¹if seen from outside (compare also [2]).

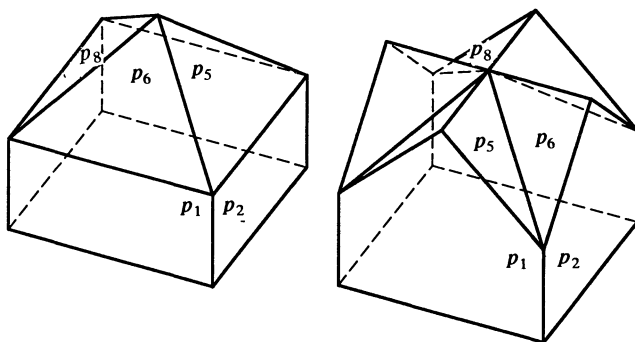


Figure 2.

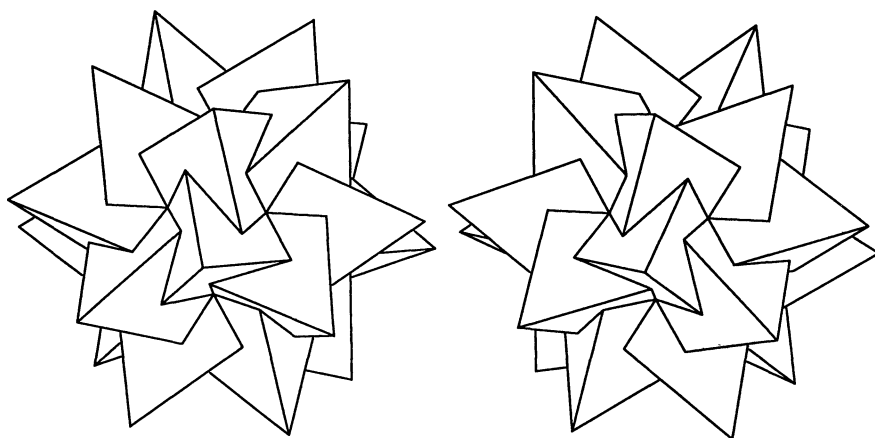


Figure 3.

(FIGURE 3, right). Then S and S' are well known stellated icosahedra (see [1], p. 5, Ef_1 and Ef_2 ; plate XVI shows the latter; the former is its reflected image).

Each two tetrahedra L_i and R_j , $i \neq j$, share exactly one vertex of D . We denote this vertex by v_{ij} . This notation has the following advantage: In order to get any transformation of the group \mathcal{S} of direct symmetries of D take any even permutation π of elements $\{1, 2, 3, 4, 5\}$. Then the corresponding isometry maps the vertex v_{ij} to $v_{\pi(i)\pi(j)}$. The points v_{ij} and v_{ji} are opposite vertices of D .

The faces of L_i and R_j opposite to v_{ij} belong to the same plane p_{ij} . The face of L_1 in plane p_{12} is a regular triangle Λ_1 with vertices v_{13}, v_{14}, v_{15} (FIGURE 4). It can be carried into the face P_2 of R_2 with vertices v_{52}, v_{32}, v_{42} by a rotation about the common center. The rotation angle

$$\varphi = \arccos \frac{3\sqrt{5} - 1}{8} = 44,4775 \dots^\circ$$

is deduced from the fact that the smallest distance between vertices of Λ_1 and P_2 equals the edge length of D .

The tetrahedra L_2 and R_2 form a stella octangula. Therefore, L_2 intersects plane p_{12} along the triangle Λ_2 of midpoints of R_2 . Λ_2 is bounded by the traces of planes p_{23}, p_{24}, p_{25} . The interior of Λ_2 is dotted in FIGURE 4.

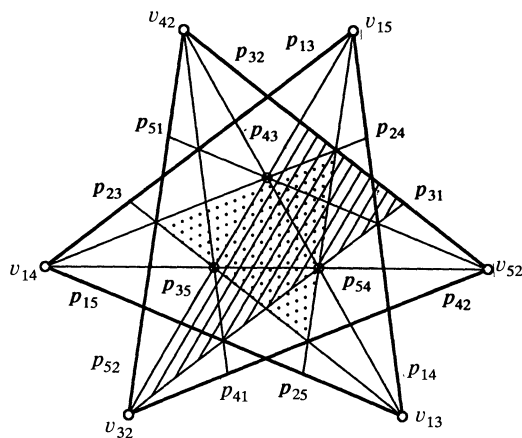


Figure 4.

L_3 intersects p_{12} in a triangle Λ_3 (hatched in FIGURE 4). The sides of Λ_3 are located on the traces of planes p_{31}, p_{32}, p_{35} . The plane p_{32} contains the vertices v_{42} and v_{52} of P_2 and therefore also a vertex of Λ_2 . The plane p_{35} passes through v_{32} and v_{15} . The plane p_{31} belongs to R_1 ; therefore its trace in p_{12} connects the midpoints of two sides of L_1 . v_{32} must be a vertex of Λ_3 .

We get the intersections Λ_4, Λ_5 of L_4, L_5 in p_{12} by rotating Λ_3 about 120° and 240° . The union $\Lambda := \Lambda_3 \cup \Lambda_4 \cup \Lambda_5$ is a superset of Λ_2 . This is a consequence of the fact that all planes p_{ij} bound a regular icosahedron polar to D . Whenever five vertices belong to the same face of D then there is a subgroup of \mathcal{S} of order 5 preserving this face. Then the corresponding planes meet at a common point on the axis of rotation. E.g. the vertices $v_{12}, v_{25}, v_{54}, v_{43}, v_{31}$ permute under the group given by the cycle (12543). Hence the planes $p_{25}, p_{54}, p_{43}, p_{31}$ have concurrent traces in plane p_{12} .

The faces of $S = L_1 \cup L_2 \cup L_3 \cup L_4 \cup L_5$ are located in the planes p_{ij} . E.g. the face of S in p_{12} is given by $\Lambda_1 \setminus \Lambda$ (FIGURE 5, left). It consists of three pentagons (compare [1], Fig. 3., where the regions 5, 6, 7, 9, 10 and 5, 6, 7, 9, 10 give FIGURE 5, upside down). The plane p_{12} contains also a face of the union S' of all right tetrahedra (FIGURE 5, right). A reflection that commutes Λ_1 and P_2 in p_{12} interchanges also the faces of S and S' . This reveals that these faces share all those

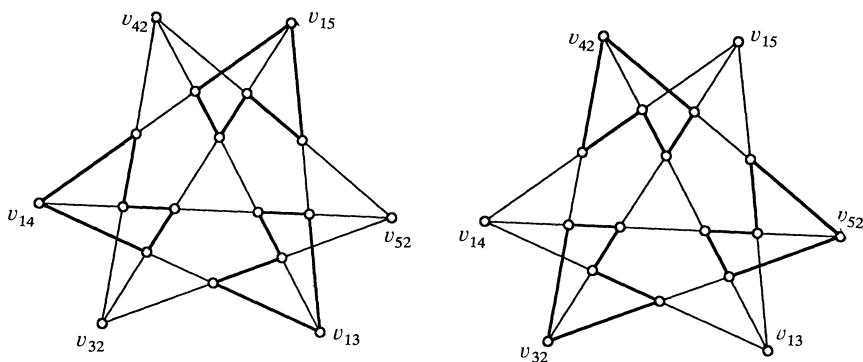


Figure 5.

vertices that are not vertices of \mathbf{D} . On the other hand all vertices of \mathbf{D} are vertices of \mathbf{S} and \mathbf{S}' too. This proves $P = P'$ and $V = V'$ for \mathbf{S} and \mathbf{S}' . We summarize as follows:

Lemma. *For the identity of two polyhedra in E^3 it is not sufficient to require equal sets of vertices and equal sets of bounding planes.*

REFERENCES

1. H. S. M. Coxeter, P. Du Val, H. T. Flather and J. F. Petrie, *The Fifty-Nine Icosahedra*, Springer-Verlag, New York Heidelberg Berlin, 1982.
2. H. Stachel, Zwei bemerkenswerte bewegliche Strukturen, *J. Geom.* 43 (1992), 14–21.

*Institute of Geometry
Technical University Vienna
A 1040 Vienna (Europe)
stachel@egmvs2.una.ac.at*

More on the Pompeiu Problem

David C. Ullrich

1. INTRODUCTION. A recent Monthly article by H. Lacquer [HL] begins with the following question:

Question. *Suppose that f is a continuous function on S and the integral of f over every α -sector vanishes, for some $\alpha \in (0, \pi)$. Must f be identically zero?*

(Here $S \subset \mathbb{R}^3$ is the two-sphere and an α -sector is the region on S bounded by two great circles meeting at an angle α .)

Theorem 1 of [HL] states that the answer is “yes” for *continuously differentiable* functions f , but the general case is left open; we shall see that the general case actually follows from the theorem in [HL], by a “regularization” argument. (In fact the hypotheses on f can be weakened still further; see the note at the end of section 3.)

One may regard the result as a sort of “tauberian theorem”. Lacking space to explain precisely what is meant by that, let us just point out that the answer to the question changes if other regions are substituted for α -sectors. For example, if we consider the same question with spherical caps instead of α -sectors then “Ungar’s Freak Theorem” states that the answer is sometimes yes and sometimes no, depending on the “radius” of the cap; see [HL] for more discussion. Some insight into what’s really going on here may be obtained by considering an analogous lower-dimensional question:

Easier Question. Suppose f is a continuous 2π -periodic function on the line, $0 < \alpha < 2\pi$, and

$$\int_x^{x+\alpha} f(t) dt = 0$$

for all $x \in \mathbb{R}$. Does this imply that f is identically zero?

Easy Answer. “Yes” if α/π is irrational, “no” otherwise.

Sketch of Easy Proof. The hypothesis says precisely that the convolution $f * \chi$ is identically zero, where χ is the 2π -periodic function taking the value 1 on the interval $[0, \alpha)$ and 0 on $[\alpha, 2\pi)$. Thus the Fourier coefficients $(f * \chi)^\wedge(n)$ must vanish for all $n \in \mathbb{Z}$.

But $(f * \chi)^\wedge(n) = \hat{f}(n)\hat{\chi}(n)$, and one easily calculates that $\hat{\chi}(0) = \alpha/2\pi$ and $\hat{\chi}(n) = (2\pi in)^{-1}(1 - e^{-in\alpha})$ for $n \neq 0$. Now if α/π is irrational this shows that $\hat{f}(n) = 0$ for all $n \in \mathbb{Z}$, which implies that f vanishes identically. On the other hand, if $\alpha = 2\pi m/n$ then $f(t) = e^{int}$ (or $f(t) = \cos(nt)$) gives the required counterexample.

(One can give a more elementary proof of the above result, without involving any Fourier analysis; the proof above is precisely analogous to the proof of the “main computational lemma” in [HL], with the circle $\mathbb{R}/2\pi\mathbb{Z}$ in place of the group $SO(3)$. It seems possible that [HL, Theorem 1] itself could be proved directly by a similar argument, but (at least) some details would need to be changed, because α -sectors are not hemispheres; in particular, there is no natural way to associate a unique α -sector to each point of the sphere.)

2. INTEGRATION ON COMPACT GROUPS. It seems like a good idea to say just a little bit about integration on compact groups before proceeding. The reader might find a proof of everything in this section in [FO], Chapters 1, 2, 7, and 10.2.

If K is a compact topological space the notation $C(K)$ will refer to the space of all (real-valued) continuous functions on K . Suppose that G is a compact group; that is, G is simultaneously a group and a compact topological space, in such a way that the group operations are continuous. A *linear functional* on $C(G)$ is a linear map $\Lambda: C(G) \rightarrow \mathbb{R}$; we say that Λ is *positive* if $\Lambda f \geq 0$ whenever $f \geq 0$ on G . It turns out that there exists a unique positive linear functional Λ on $C(G)$ satisfying the following two conditions:

$$\Lambda \mathbf{1} = 1, \tag{2.1}$$

$$\Lambda g = \Lambda f \quad \text{if} \quad g(x) \equiv f(x_0 x) \quad \text{for some } x_0 \in G. \tag{2.2}$$

(Here “ $\mathbf{1}$ ” denotes the element of $C(G)$ with $\mathbf{1}(x) = 1$ for all $x \in G$.) The existence of a functional Λ satisfying all these properties is not obvious or trivial, but assuming that such a thing exists the reader may easily derive the following two inequalities:

$$\text{If } f, g \in C(G) \quad \text{and} \quad f \leq g \quad \text{then} \quad \Lambda f \leq \Lambda g. \tag{2.3}$$

$$\text{If } g = |f| \quad \text{then} \quad |\Lambda f| \leq \Lambda g. \tag{2.4}$$

The value of Λf is usually denoted “ $\int_G f(x) dx$ ” and is known as the “integral” of f with respect to “Haar measure”.

Of course given a compact group G one can often find explicit formulas for $\int_G f(x) dx$. The case of interest below will be $G = SO(3)$, the group of (real) 3×3 matrices with determinant one. Perhaps the most transparent way to give a

formula for the invariant integral in this case is as follows: For $\zeta \in S$ and $t \in \mathbb{R}$ let $R(\zeta, t)$ denote a counter-clockwise rotation through an angle t about the axis ζ . It is not hard to see that $SO(3) = \{R(\zeta, t): \zeta \in S, t \in [0, 2\pi)\}$; in fact if $T \in SO(3)$, $T \neq I$, then $T = R(\zeta, t)$ ($0 < t < 2\pi$) in precisely two ways, and

$$\int_{SO(3)} f(T) dT = \frac{1}{4\pi} \int_S \frac{1}{2\pi} \int_0^{2\pi} f(R(\zeta, t)) dt d\sigma(\zeta).$$

(Here “ $d\sigma$ ” means “with respect to surface area”.) One may also “represent” the elements of $SO(3)$ in terms of quaternions, leading to another formula for the integral; see [ZH] pp. 33–35.

3. THE POMPEIU PROBLEM. We now suppose that $f \in C(S)$, $\alpha \in (0, \pi)$, and the integral of f over any α -sector vanishes. The plan is to find a sequence of *continuously differentiable* functions f_1, f_2, \dots with the same properties, and which also converge uniformly to f . Theorem 1 of [HL] will then show that $f_n = 0$ for each n , so that $f = 0$.

To begin, suppose that $h = f \circ T$ for some $T \in SO(3)$. It follows that the integral of h over any α -sector is zero, because surface area on S is rotation-invariant and T takes α -sectors to α -sectors: if V is an α -sector then

$$\int_V h d\sigma = \int_V f \circ T d\sigma = \int_{T(V)} f d\sigma = 0. \quad (3.1)$$

(Note that T preserves distances; hence T maps geodesics to geodesics, which is to say that it maps great circles to great circles. Since T also preserves angles it follows that $T(V)$ must be an α -sector.)

Now given a function $g \in C(SO(3))$ we define $f * g \in C(S)$ by the formula

$$f * g(\zeta) = \int_{SO(3)} f(T^{-1}\zeta) g(T) dT \quad (\zeta \in S), \quad (3.2)$$

or simply

$$f * g = \int_{SO(3)} (f \circ T^{-1}) g(T) dT. \quad (3.3)$$

It follows that the integral of $f * g$ over any α -sector is zero. Indeed, if V is an α -sector then

$$\begin{aligned} \int_V f * g(\zeta) d\sigma(\zeta) &= \int_V \left(\int_{SO(3)} f(T^{-1}\zeta) g(T) dT \right) d\sigma(\zeta) \\ &= \int_{SO(3)} g(T) \left(\int_V f(T^{-1}\zeta) d\sigma(\zeta) \right) dT = \int_{SO(3)} 0 dT = 0, \end{aligned} \quad (3.4)$$

by (3.1). (The technical conditions needed to justify the exchange of the order of integration are easily verified here; see [FO], Chapter 2.5.)

Now define a metric on $SO(3)$ by setting $d(T_1, T_2) = \sup_{\zeta \in S} \|T_1(\zeta) - T_2(\zeta)\|$, and choose a sequence of *continuously differentiable* functions $g_n \in C(SO(3))$ ($n = 1, 2, \dots$) such that

$$g_n \geq 0, \quad (3.5)$$

$$\int_{SO(3)} g_n(T) dT = 1, \quad (3.6)$$

and

$$g_n(T) = 0 \quad (T \in SO(3), d(T, I) \geq 1/n). \quad (3.7)$$

Set $f_n = g_n * f$.

The fact that g_n is continuously differentiable implies that f_n is continuously differentiable. Formula (3.4) shows that the integral of f_n over any α -sector is zero, and hence Theorem 1 of [HL] shows that $f_n = 0$; we need only show that $f_n \rightarrow f$ as $n \rightarrow \infty$.

So we suppose $\varepsilon > 0$, and we choose N so that $|f(\zeta) - f(\xi)| < \varepsilon$ whenever $\|\zeta - \xi\| < 1/N$ ($\zeta, \xi \in S$). A few standard tricks show that $|f(\zeta) - f_n(\zeta)| < \varepsilon$ for all $\zeta \in S$ if $n > N$: First, note that

$$\begin{aligned} f(\zeta) - f_n(\zeta) &= f(\zeta) - \int_{SO(3)} f(T^{-1}\zeta) g_n(T) dT \\ &= \int_{SO(3)} (f(\zeta) - f(T^{-1}\zeta)) g_n(T) dT; \end{aligned}$$

here we have used nothing but (3.6) and linearity. It follows from (2.4) that

$$|f(\zeta) - f_n(\zeta)| \leq \int_{SO(3)} |f(\zeta) - f(T^{-1}\zeta)| g_n(T) dT.$$

Now if $n > N$ then $|f(\zeta) - f_n(\zeta)| g_n(T) \leq \varepsilon g_n(T)$ for all $\zeta \in S$ and $T \in SO(3)$ (consider separately the cases $d(T, I) < 1/n$ and $d(T, I) \geq 1/n$), so that

$$|f(\zeta) - f_n(\zeta)| \leq \varepsilon \int_{SO(3)} g_n(T) dT = \varepsilon \quad (\zeta \in S, n > N),$$

by (2.3). Hence $f_n \rightarrow f$ uniformly on S , as claimed.

Note. If the reader knows some real analysis (s)he may use essentially the same argument to show that if f is a Lebesgue-integrable function on the sphere and the integral of f over every α -sector vanishes then $f = 0$ almost everywhere: The fact that $C(S)$ is dense in $L^1(S)$ shows that $f_n \rightarrow f$ in $L^1(S)$. (The novice should perhaps be warned that it is somewhat more difficult to show that $f_n \rightarrow f$ almost everywhere; in fact this is false unless one is a bit more careful in the choice of g_n .)

REFERENCES

-
- [FO] Folland, G. B. *Real Analysis*. John Wiley & Sons, New York (1984).
 [HL] Lacquer, H. T., "The Pompeiu Problem", *Amer. Math. Monthly* **100** (1993) #5, 461–467.
 [ZH] Zhelobenko, D. P., *Compact Lie Groups and Their Representations*. AMS Translations Volume 40, 1973.

Department of Mathematics
Oklahoma State University
Stillwater, OK 74078
ullrich@math.okstate.edu

UNSOLVED PROBLEMS

Edited by: **Richard Guy**

In this department the MONTHLY presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial or related results. Typescripts should be sent to Richard Guy, Department of Mathematics & Statistics, The University of Calgary, Alberta, Canada T2N 1N4.

Every Number is Expressible as the Sum of How Many Polygonal Numbers?

Richard K. Guy

Sums of squares have been extensively studied, and the following three theorems, due essentially to Lagrange, Legendre and Fermat, are among the most famous in number theory. The first was stated in 1621 by Bachet.

Every number is expressible as the sum of **four** squares.

The numbers expressible as the sum of **three** squares are just those **not** of the form $4^k(8l + 7)$, $k, l \geq 0$.

The numbers expressible as the sum of **two** squares are just those whose “bad” prime factors (those of shape $4k - 1$) occur to an **even** power.

Fermat is perhaps most famous for the fact that the margin of his copy of Diophantus wasn’t wide enough to hold a certain proof, but another claim of his, that every number is expressible as the sum of k k -gonal numbers, is worth noting. He wrote to Pascal on 1654-09-25:

“Ce que vous y trouverez de plus important regarde la proposition que tout nombre est composé d’un, de deux ou de trois triangles; d’un, de deux, de trois ou de quatre carrés; d’un, de deux, de trois, de quatre ou de cinq pentagones; d’un, de deux, de trois, de quatre, de cinq ou de six hexagones, et à l’infini.

Pour y parvenir, il faut démontrer que tout nombre premier, qui surpasse de l’unité un multiple de 4, est composé de deux carrés, comme 5, 13, 17, 29, 37, etc.

He promised to devote an entire book to this subject, but it didn’t appear, and we had to wait for Cauchy [2] to give the first proof. Meanwhile, for triangular numbers, there was Gauss’s famous 1796-07-10 diary entry (see [1]):

$$\text{ETPHKA!} \quad \text{num} = \Delta + \Delta + \Delta,$$

i.e., every number is expressible as the sum of three triangular numbers.

Squares and triangles are the most familiar members of the family of **polygonal numbers**. The r -th k -gonal number is

$$\frac{1}{2}r((k-2)r - (k-4)).$$

Squares and triangles have the property that we obtain no new numbers when we throw in those of negative rank, $r < 0$.

$$k = 4: (-r)^2 = r^2 \quad k = 3: \frac{1}{2}(-r)(-r+1) = \frac{1}{2}(r-1)((r-1)+1).$$

But when we come to the pentagonal and other k -gonal numbers for $k \geq 5$ (and $k \leq -1$) we get a different set if we allow numbers of negative rank.

TABLE OF k -GONAL NUMBERS

-90	-65	-44	-27	-14	-5	0	1	-2	-9	-20	-35	-54	-77	-104	-135	-170
-69	-50	-34	-21	-11	-4	0	1	-1	-6	-14	-25	-39	-56	-76	-99	-125
-48	-35	-24	-15	-8	-3	0	1	0	-3	-8	-15	-24	-35	-48	-63	-80
-27	-20	-14	-9	-5	-2	0	1	1	0	-2	-5	-9	-14	-20	-27	-35
-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10
15	10	6	3	1	0	0	1	3	6	10	15	21	28	36	45	55
36	25	16	9	4	1	0	1	4	9	16	25	36	49	64	81	100
57	40	26	15	7	2	0	1	5	12	22	35	51	70	92	117	145
78	55	36	21	10	3	0	1	6	15	28	45	66	91	120	153	190
99	70	46	27	13	4	0	1	7	18	34	55	81	112	148	189	235
120	85	56	33	16	5	0	1	8	21	40	65	96	133	176	225	280

In the table, the values of r and k are the **bold row** and **kolumn**. The $(-k)$ -gonal numbers are 1 minus the $(4+k)$ -gonal numbers: more precisely, the r -th $(-k)$ -gonal number is 1 minus the $(1-r)$ -th $(4+k)$ -gonal number.

Let us jump to the hexagonal numbers. It is well known that every hexagonal number is triangular:

$$r(2r-1) = \frac{1}{2}(2r-1)((2r-1)+1)$$

but not so well known that the converse is also true:

$$\frac{1}{2}r(r+1) = \left(\frac{r+1}{2}\right)\left(2\left(\frac{r+1}{2}\right)-1\right) \text{ or } \left(\frac{-r}{2}\right)\left(2\left(\frac{-r}{2}\right)-1\right)$$

according as r is odd or even. With this interpretation, it follows from Gauss's result that

Every number is the sum of three hexagonal numbers.

But if we restrict ourselves to hexagonal numbers of positive rank and put

$$n = r_1(2r_1-1) + r_2(2r_2-1) + r_3(2r_3-1),$$

$$8n+3 = (4r_1-1)^2 + (4r_2-1)^2 + (4r_3-1)^2$$

then, although every number $8n+3$ is expressible as the sum of three odd squares, they will not necessarily be squares of numbers of shape $4r-1$ with $r > 0$. Many numbers (what fraction of the whole?) require four hexagonal numbers of positive rank; several, e.g.,

$$5, 10, 20, 25, 38, 39, 54, 65, 70, 114, 130, \dots,$$

require five, and 11 and 26 require six. Which numbers require five? Are there any others, or is every sufficiently large number the sum of four hexagonal numbers of nonnegative rank? Perhaps every sufficiently large number is the sum of only three such numbers?

For pentagonal numbers, if

$$n = \frac{1}{2}r_1(3r_1 - 1) + \frac{1}{2}r_2(3r_2 - 1) + \frac{1}{2}r_3(3r_3 - 1),$$

$$24n + 3 = (6r_1 - 1)^2 + (6r_2 - 1)^2 + (6r_3 - 1)^2$$

and every such number is expressible as the sum of three odd squares. Moreover, either none or all of these squares are multiples of 3. If the latter, then n must have been of shape $3m + 1$ and, after removal of a factor 9, $8m + 3$ is the sum of three odd squares. Repeating the process if necessary, we may assume that the squares are of shape $6r \pm 1$. So, if negative ranks are allowed, every positive integer can be expressed as the sum of three pentagonal numbers. But, restricting themselves to pentagonal numbers of positive rank, Richard Blecksmith & John Selfridge found six numbers among the first million, namely

$$9, 21, 31, 43, 55 \text{ and } 89,$$

which require five summands, and two hundred and four others, the largest of which is 33066, which require four.

They believe that they have found them all.

Can all sufficiently large numbers be expressed as the sum of three pentagonal numbers of nonnegative rank? Equivalently, is every sufficiently large number of shape $24n + 3$ expressible as the sum of three squares of numbers of shape $6r - 1$?

For heptagonal numbers, $\frac{1}{2}r(5r - 3)$, even if we allow those of negative rank, the situation is less clear. If

$$n = \frac{1}{2}r_1(5r_1 - 3) + \frac{1}{2}r_2(5r_2 - 3) + \frac{1}{2}r_3(5r_3 - 3),$$

$$40n + 27 = (10r_1 - 3)^2 + (10r_2 - 3)^2 + (10r_3 - 3)^2$$

and although every number $40n + 27$ is expressible as the sum of three squares, these will not necessarily be of shape $(10r \pm 3)^2$. For example, 427 is uniquely expressible as $9^2 + 11^2 + 15^2$ and 667 only as $1^2 + 15^2 + 21^2$ or $9^2 + 15^2 + 19^2$, so 10 and 16 are not the sum of three heptagons. Nor is 76; are there others?

If n is the sum of three octagonal numbers,

$$n = r_1(3r_1 - 2) + r_2(3r_2 - 2) + r_3(3r_3 - 2)$$

$$3n + 3 = (3r_1 - 1)^2 + (3r_2 - 1)^2 + (3r_3 - 1)^2$$

there will be no such representation if $3n + 3 = 4^k(8l + 7)$, i.e., if $n = 4^k(8l + 5) - 1$, e.g., $n = 8l + 4$, $32l + 19$, $128l + 79, \dots$ but otherwise it seems that there is always a representation, provided we allow numbers of both positive and negative rank. If so, this implies that all multiples of 3, other than those of shape $4^k(24l + 15)$ are representable as the sum of three squares, none of them being multiples of 3.

It is clear that, for $k \geq 8$, $k - 4$ requires $k - 4$ ones to represent it as the sum of k -gonal numbers, while $2k - 1$ and $5k - 4$ require k k -gonal numbers of positive rank to represent them (k and $k - 1$ ones; four k s and $k - 4$ ones). But, for n sufficiently large, presumably a much smaller number suffices.

What theorems are there, stating that all numbers of a suitable shape are expressible as the sum of three (say) squares of numbers of a given shape?

We can also ask for the **number of representations** of n as the sum of polygonal numbers. Jacobi (see [3, 5, 7, 8]) gave the answers for two or four squares:

The number of representations of n as the sum of two squares is $4\{d_1(n) - d_3(n)\}$, where $d_i(n)$ is the number of divisors of n which are $\equiv i \pmod{4}$.

The number of representations of n as the sum of four squares is $8(2 + (-1)^n)\sigma_d(n)$, where $\sigma_d(n)$ is the sum of the odd divisors of n .

This last number can also be read as “8 times the sum of those divisors of n which are not multiples of 4.” Notice that these results count representations as different if they differ only in order, or in sign. For example $4 = (\pm 2)^2 + 0^2 = 0^2 + (\pm 2)^2$ has $2 + 2 = 4$ representations and $4 = (\pm 1)^2 + (\pm 1)^2 + (\pm 1)^2 + (\pm 1)^2 = (\pm 2)^2 + 0^2 + 0^2 + 0^2$ has $2^4 + 4 \cdot 2 = 24$.

For the corresponding result for two triangular numbers, see [5 & 6]; the number of representations of n as the sum of two triangular numbers is $d_1(4n + 1) - d_3(4n + 1)$.

Can we find corresponding results for any of the other polygonal numbers?

REFERENCES

1. G. E. Andrews, ETPHKA! num = $\Delta + \Delta + \Delta$, *J. Number Theory*, 23 (1986) 285–293.
2. A. Cauchy, *Mém. Sci. Math. et Phys. l'Inst. France* (1), 14 (1813–15) 177–220.
3. John A. Ewell, A simple derivation of Jacobi's four-square formula, *Proc. Amer. Math. Soc.*, 85 (1982) 323–326.
4. John A. Ewell, A simple proof of Fermat's two-square theorem, this MONTHLY, 90 (1983) 635–637.
5. John A. Ewell, On sums of triangular numbers and sums of squares, this MONTHLY, 99 (1992) 752–757.
6. John A. Ewell, On representations of numbers by sums of two triangular numbers, *Fibonacci Quart.*, 30 (1992) 175–178.
7. M. D. Hirschhorn, A simple proof of Jacobi's four-square theorem, *J. Austral. Math. Soc. Ser. A*, 32 (1982) 61–67.
8. M. D. Hirschhorn, A simple proof of Jacobi's two-square theorem, this MONTHLY, 92 (1985) 579–580.

*Department of Mathematics and Statistics
The University of Calgary
Calgary, Alberta
Canada T2N 1N4*

Excerpt from *The Long Comeback of Henry Roth: Call It Miraculous* by Leonard Michaels, The New York Times Book Review, August 15, 1993.

In one of Walton's letters, written to a friend in 1960, when Mr. Roth was in Maine, she says, “He cannot write, and calms himself and his depressions, which seem many, by studying not Greek, as with me, but math.”

Mr. Roth says he used the textbook “Calculus and Analytic Geometry,” by George Thomas of M.I.T. He even sent Professor Thomas a letter “very respectfully” correcting some of the solutions (“Who else would be crazy enough to do every problem in the book?” he says in his self-deprecating way.) To think of the self-taught Mr. Roth doing calculus problems reveals much about his exertion, in the freezing and scary darkness of Maine, not to write fiction.

[Author's answer to his question: Donald Knuth—see his interview in *Mathematical People*, p. 185–186.]

Contributed by Stephen B. Maurer
Department of Mathematics
Swarthmore College
Swarthmore, PA 19081-1397

ANN ELIZABETH HIRST is a graduate of the universities of London and Southampton in the UK. She has taught Mathematics in schools, a college of Art, the Open University and the University of Southampton, and her students' ages have ranged from 11 to 75. Her main interests lie in Geometry (particularly in the Mathematics of Surfaces) and Mathematics Education; (she was joint editor of the *Proceedings of ICME-6*). A large proportion of her spare time is spent in choral singing or listening to music.

RUTHERFORD ARIS was introduced to mathematical modeling by apprenticeship to C. H. Bosanquet in the Research Department of I.C.I. at Billingham-on-Tees. Under the eye of this brilliant and eccentric man he tackled a number of problems ranging from the anisotropy of fire-bricks to the efficiency of venturi scrubbers. In 1955–56 he spent a year working with N. R. Amundson, finding Hopf and homoclinic bifurcations in the control of a stirred tank. After two years of teaching mathematics at Edinburgh, he returned to the chemical engineering department at the University of Minnesota and there he has been ever since. He is a Fellow of the American Academy of Arts and Sciences and of the Institute of Mathematics and its Applications, a Member of the National Academy of Engineering and, among various others, recipient of the Richard Bellman Control Heritage Award.

DAVID L. WEBB received his undergraduate training at the University of Tennessee and his Ph.D. from Cornell University under the direction of Kenneth S. Brown. Following a year as a Postdoctoral Fellow at the University of Waterloo, he was Assistant Professor at Washington University from 1984–1990 and Associate Professor until 1992, when he joined the faculty of Dartmouth College. He was recently awarded an AMS Centennial Research Fellowship. His research interests include algebraic K -theory and differential geometry.

Letter to the Editor

Recently Norbert Hegyvári [1] elegantly subsumed proofs of the irrationality of the fraction consisting of the decimal representations of consecutive primes by showing the following:

Let a_1, a_2, \dots be a sequence of distinct positive integers with decimal representations $(a_1), (a_2), \dots$. Suppose the fraction $\alpha = 0.(a_1)(a_2)\dots$ is rational. Then $\sum 1/a_i$ converges.

The proof, however, can be considerably simplified. Omitting a finite number of terms of the sequence if necessary, we may assume α has a periodic decimal representation. Call the period s . For a fixed k consider the terms a_i for which $10^{k-1} \leq a_i < 10^k$, i.e. those with exactly k digits. As a term is uniquely defined by its number of digits and its starting location within α , there can be no more than s terms of k digits. Hence, $\sum 1/a_i \leq \sum s/10^{k-1}$. The latter series converges and the result follows.

[1] N. Hegyvári, On some irrational decimal fractions, *Amer. Math. Monthly* 100 (1993), 779–780.

Laird E. Taylor
Mathematics Department
California State University, Bakersfield
9001 Stockdale Highway
Bakersfield, CA 93311-1099
larry@ultrix4.csuak.edu

PROBLEMS AND SOLUTIONS

Edited by:

Richard T. Bumby, Fred Kochman and Douglas B. West

Proposed problems should be sent to the MONTHLY PROBLEMS address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.

Solutions of published problems should arrive before July 31, 1994 at the MONTHLY PROBLEMS address given on the inside front cover. Solutions should be typed with double spacing, including the problem number and the solver's name and mailing address. Two copies suffice. A self-addressed postcard or label should be included if an acknowledgment is desired.

*An asterisk (*) after the number of a problem, or part of a problem, indicates that no solution is currently available. Partial solutions will be useful in such cases. Otherwise, the published solution is likely to be based on a solution which is complete and correct. Of course, an elegant partial solution or a method leading to a more general result is always useful and welcome. In addition, references to other appearances of MONTHLY problems or to solutions of these problems in the literature are also solicited.*

PROBLEMS

10361. *Proposed by Emil A. Cornea, University of Bucharest, Bucharest, Romania, and Florin N. Diacu, University of Victoria, Victoria, B.C., Canada.*

Do there exist nonlinear C^1 functions $f: \mathbb{R} \rightarrow \mathbb{R}$ such that for any rational x , $f(x)$ is also rational and for any irrational x , $f(x)$ is also irrational?

10362. *Proposed by Hans Liebeck and Anthony Osborne, University of Keele, England.*

Let A be a real orthogonal matrix without eigenvalue 1. Let B be obtained from A by replacing one of its rows or one of its columns by its negative. Show that B has 1 as an eigenvalue.

10363. *Proposed by Joseph M. Santmyer, California University of Pennsylvania, California, PA.*

If M, N are integers satisfying $1 \leq m \leq n - 1$, prove that

$$\binom{2n - m - 1}{2n - 2m - 1} - \binom{n - 1}{m} = \sum_k \sum_j \binom{k + j}{k} \binom{2n - m - 2k - j - 3}{2(n - m - k - 1)}.$$

10364. *Proposed by Frank Schmidt, Arlington, VA.*

Let S_{2n} denote the symmetric group of degree $2n$. Let E_{2n} (respectively O_{2n}) be the set of those permutations in S_{2n} all of whose cycle lengths are even (respectively odd). Show that E_{2n} and O_{2n} are equinumerous by finding an explicit bijection between them.

10365. *Proposed by Daniel Tisdale (student), University of Maryland, Baltimore, MD.*

Consider the square $\{(x, y): 0 \leq x, y \leq 1\}$ divided into n^2 equal small squares by the lines $x = i/n$ and $y = j/n$. For $1 \leq i \leq n$, let $x_i = i/n$ and

$$d_i = \min_{0 \leq j \leq n} \left| \sqrt{1 - x_i^2} - \frac{j}{n} \right|.$$

Determine

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n d_i,$$

or show that it does not exist.

10366. *Proposed by Raphael M. Robinson, University of California, Berkeley, CA.*

Let \mathbf{C} denote the Cantor set. If s is a real number, by $\mathbf{C} + s$ is meant the set of all sums $c + s$ where $c \in \mathbf{C}$. Find a value of s such that $\mathbf{C} + s$ contains no rational number. This value of s must be given explicitly; an existence proof does not suffice.

10367. *Proposed by Donald R. Chalice, Western Washington University, Bellingham, WA.*

Let \mathbf{C} be the Cantor set in $[0, 1]$, and let \mathbf{E} be the set of endpoints of the removed intervals (together with 0 and 1). Let $\mathbf{F} = \mathbf{C} - \mathbf{E}$ and let p be the point $(1/2, 1/2)$. For any $c \in \mathbf{C}$, let L_c be the line segment from p to c , let Q_c be the points on L_c with rational ordinates and I_c the points of L_c with irrational ordinates. The set

$$\mathbf{T} = \left(\bigcup_{e \in \mathbf{E}} Q_e \right) \cup \left(\bigcup_{f \in \mathbf{F}} I_f \right)$$

has the property that \mathbf{T} is connected, but $\mathbf{T} - p$ is totally disconnected. Consider instead

$$\mathbf{T}_0 = \left(\bigcup_{e \in \mathbf{E}} I_e \right) \cup \left(\bigcup_{f \in \mathbf{F}} Q_f \right)$$

obtained by interchanging the roles of points with rational and irrational ordinates. Is \mathbf{T}_0 connected?

NOTES

Notes: (10363) No limits of summation are included in the statement since the sums may be extended over all integers with the convention that $\binom{n}{k} = 0$ unless $0 \leq k \leq n$. For the convenience of readers, we note that, in this case, the nonzero terms occur for $0 \leq k \leq n - m - 1$ and $0 \leq j \leq m - 1$. (10364) It is known that $\#(E_{2n}) = \#(O_{2n}) = ((2n - 1)(2n - 3) \cdots 3 \cdot 1)^2$, but proofs of this are based on separate counting of the two sets. (10365) This is a representative of a large family of questions. The d_i are the shortest distances on the vertical lines between grid points and the graph of the unit circle in the first quadrant; and one could also consider the behavior of the *signed distance* from the circle to the closest grid point. The sum of absolute values is only one way to describe the n -tuple of distances. The pattern of signed distances for $n = 200$ or so illustrates the difficulty of representing smooth curves by grid points. (10367) The set T was constructed by B. Knaster and K. Kuratowski in the paper “Sur les ensembles connexes”, *Fund. Math.* 2 (1921), 206–255 (This can also be found in L. A. Steen & J. A. Seebach, Jr., *Counterexamples in Topology*, p. 145). The set T_0 appears in B. Gelbaum and J. Olmstead, *Counterexamples in Analysis*, p. 146 in place of the original example of Knaster and Kuratowski. Thus we are asking whether this example has the advertised property.

SOLUTIONS

A Maximal Average-Excluding Set

E 3474 [1991, 956]. *Proposed by Andrew Lenard, Indiana University, Bloomington, IN.*

Let S be the set of nonnegative real numbers whose expansion to the base 4 involves only the digits 0 and 1.

(a) Prove that if $x \in S$, $y \in S$, and $x \neq y$, then $(x + y)/2 \notin S$.

(b) Suppose T is a set of nonnegative real numbers properly containing S . Prove that there exist $x \in T$, $y \in T$, $x \neq y$ such that $(x + y)/2 \in T$.

Composite solution of part (a) by many solvers. If $(x + y)/2 \in S$, then the base 4 expansion of $x + y$ must involve only the digits 0 and 2. But since $x \neq y$, the base 4 expansion of $x + y$ must in fact include at least one 1. So $(x + y)/2 \notin S$ (since for elements of S the expansion is unambiguously determined).

Solution of part (b) by NSA Problems Group, Fort Meade, MD. It suffices to show that for any nonnegative $x \notin S$, there are $y, z \in S$ with $x + y = 2z$. Sup-

pose we write $x = \sum_{i \leq N} x_i 4^i$, $x_i = 0, 1, 2$, or 3 . Let $t = \sum_{i \leq N+1} 4^i$ and define integers u_i by the base 4 expansion of $u = t + x = \sum_{i \leq N+1} u_i 4^i$. Next, for each integer u_i define integers y_i and z_i with values given by the following table:

u_i	y_i	z_i
0	1	0
1	0	0
2	1	1
3	0	1

Then define $y, z \in S$ by

$$y = \sum_{i \leq N+1} y_i 4^i$$

and

$$z = \sum_{i \leq N+1} z_i 4^i.$$

Since each row of the table satisfies $2z_i + 1 = u_i + y_i$, we find that

$$2z - y = \sum_{i \leq N+1} (2z_i - y_i) 4^i = \sum_{i \leq N+1} (u_i - 1) 4^i = u - t = x$$

as desired. Note that if the construction is applied to $x \in S$, then since each u_i is 1 or 2 we end up with $y = z$ and hence $x = y$, in conformity with part a.

Editorial comment. For part (b), most respondents took an approach requiring a slightly awkward analysis of the propagation of carries in the sum of base 4 numbers $x + y$ for $y \in S$, $x \notin S$. In about half the responses the presentation was significantly less clear than the selected solution. The device adopted above of introducing the number t allowed the carries to take care of themselves. Jiro Fukuta obtained a similar effect by using a base 4 expansion in which the digits are $-1, 0, 1$, and 2 .

Albert Nijenhuis employed Cantor set constructions. The set A_n consists of the numbers in the interval $[0, 1]$ having a base 4 expansion whose first n digits are 0 or 1, and B_n are those whose first n digits are 0 or 2. Thus $A_1 = [0, 1/2]$ and $B_1 = [0, 1/4] \cup [1/2, 3/4]$; while the intersection of the A_n is $S \cap [0, 1)$ and the intersection of the B_n gives the set of doubles of those numbers. Given $x \in [0, 1/4]$, he studies the set of $y_n \in A_n$ for which $x + y_n \in B_n$. The digits of such y_n are constructed as in other solutions, but only n digits are required. The limit process is postponed and aided by the compactness of the sets A and B .

Raphael Robinson gave an argument that the number y constructed in part (b) is unique, unless the expansion of x terminates in an infinite sequence of 2's. For instance, if $x = .222\dots$ then the construction above yields $y = 1.1111\dots$, while in fact $y = 0$ would also work.

The proposer noted that the problem could be rephrased to assert that S is a maximal average-excluding subset of \mathbb{R} . Robert High made a similar observation, and went on to ask if there are other *closed* maximal average-excluding subsets of \mathbb{R} . There is no difficulty in obtaining maximal average-excluding sets by Zorn's lemma, but a construction of a different closed set with this property, or a proof that this property uniquely characterizes S up to affine equivalence would be of interest.

Solved also by K. F. Andersen (Canada), C. D. Ashbacher, J. W. Benham, D. M. Bloom, D. Callan, R. J. Chapman (U.K.), M. Dindos (Slovakia), J. Fujuta (Japan), R. High, K. S. Kedlaya (student),

N. Komanda, A. F. Martin, A. Nijenhuis, R. E. Prather, A. Riese, R. M. Robinson, R. L. Schilling (Germany), A. Stein, Anchorage Math Solutions Group, Western Maryland College Problems group, and the proposer. One solution of part (a) only and three incorrect solutions were received.

The Smallest Factorial That Is a Multiple of n

6674 [1991, 965]. *Proposed by Paul Erdős, Hungarian Academy of Sciences, Budapest, Hungary.*

If n is an integer greater than 1, let $P(n)$ denote the largest prime factor of n . Prove that $x|P(n)!$ for almost all n , i.e., prove that if

$$S(x) = \{n \leq x: n \nmid P(n)!\},$$

then

$$\lim_{x \rightarrow \infty} |S(x)|/x = 0.$$

Solution by Ilias Kastanas, California State University, Los Angeles, CA. Let $S'(x) = \{n \leq x: n|P(n)!\}$; we will show that $d = \lim_{x \rightarrow \infty} S'(x)/x$ is equal to 1.

Let Q denote the set of square free natural numbers, and introduce

$$Q(x) = |Q \cap \{1, 2, \dots, x\}|$$

$$Q_\alpha = |\{n: n \in Q \text{ and all prime factors of } n < \alpha\}|.$$

For any integer $k > 0$ let $f(k)$ be the least prime greater than $3k$. If $q \in Q$ and q is divisible by some prime $\geq f(k)$, then $k^2q|P(q)!$, since $2k$ and $3k$ are factors of $P(q)!/q$. So, there are at least $Q(x/k^2) - Q_{f(k)}$ integers of the form k^2q in $S'(x)$, and

$$\lim_{x \rightarrow \infty} \frac{Q(x/k^2) - Q_{f(k)}}{x} = \lim_{x \rightarrow \infty} \frac{Q(x/k^2)}{x} = \frac{1}{\zeta(2)} \frac{1}{k^2}.$$

For the last equation, see G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford, 1980, Theorem 333 in §18.5. Since the sets k^2Q are disjoint for distinct k , we have

$$1 \geq d \geq \frac{1}{\zeta(2)} \sum_{k=1}^n \frac{1}{k^2}$$

for any $n > 0$. Therefore,

$$d \geq \frac{1}{\zeta(2)} \cdot \zeta(2) = 1.$$

Editorial comment. In preparing the problem for publication, the editors obtained a solution from Kevin Ford based on a division into cases depending on the number of prime factors of n and the size of $P(n)$. He was able to show that $|S(x)| = O(x/\log x)$ by this method, and indicated that stronger results could be obtained using *sieve theoretic* estimates.

No other solutions were received.

A Characterization of Special Quadratic Irrationals

10218 [1992, 362]. *Proposed by David Dwyer, University of Evansville, Evansville, IN.*

For positive real numbers r and positive integers n , put

$$\phi(n, r) = (nr) + (\lfloor nr \rfloor r),$$

where $(x) = x - \lfloor x \rfloor$ denotes the “fractional part” of x .

Find $\{r \in \mathbb{R}^+ : \phi(n, r) > 1 \text{ for all } n \in \mathbb{Z}^+\}$.

Solution by Richard Stong, Rice University, Houston, TX. The numbers $r > 0$ for which $\phi(n, r) > 1$ for all $n \in \mathbb{Z}^+$ are exactly the numbers

$$r = (k + \sqrt{k^2 + 4m})/2$$

for integers $k \geq m \geq 1$, or equivalently the positive roots of the equations $x^2 - kx - m = 0$.

Suppose first that r is such a number. Then it is easy to verify that r is irrational, with $k < r < k + 1$, so that $\lfloor r \rfloor = k$. Since r is irrational, for any positive integer n we may write $nr = p + \theta$ for some positive integer p and $0 < \theta < 1$. Then (writing $\langle \cdot \rangle$ for fractional part), $\langle nr \rangle = \theta$ and $\lfloor nr \rfloor r = pr = (nr - \theta)r = nr^2 - \theta r = n(kr + m) - \theta r = kp + k\theta + nm - \theta r = kp + nm + (k - r)\theta$. Since $0 > k - r > -1$ we see that

$$\langle \lfloor nr \rfloor r \rangle = 1 - (r - k)\theta,$$

$$\phi(n, r) = 1 + (k + 1 - r)\theta > 1.$$

Thus, all numbers r of the stated form work.

Suppose r is not of the stated form. If r is a rational number p/q then one easily checks that $\phi(q, r) < 1$. Thus we may assume r is irrational. Let $k = \lfloor r \rfloor$ and set $s = r^2 - kr = r(r - k)$. Since $k < r < k + 1$ we necessarily have $0 < s < k + 1$. Therefore, since r is not of the specified form, $s = r^2 - kr$ cannot be an integer.

Claim. *There is a positive integer n with*

$$1 - \langle nr \rangle \geq \langle ns \rangle \geq \langle nr \rangle.$$

Granting the claim and using the specified n , write

$$nr = p + \theta \quad \text{and} \quad ns = q + \sigma$$

for non-negative integers p, q , and with $0 < \theta \leq \sigma \leq 1 - \theta$. Then $\langle nr \rangle = \theta$ and $\lfloor nr \rfloor r = pr = nr^2 - r\theta = nkr + ns - r\theta = pk + q + \sigma + (k - r)\theta$. Since $0 > k - r > -1$ and $\sigma \geq \theta$ we see that $\langle \lfloor nr \rfloor r \rangle = \sigma - (r - k)\theta$, and from $\sigma \leq 1 - \theta$ we find $\phi(n, r) = \sigma + (k + 1 - r)\theta < 1$, as required.

To establish the claim, consider the set $A = \{(nr, ns) | n \in \mathbb{Z}^+\}$ as elements of the torus $T = \mathbb{R}^2/\mathbb{Z}^2$. The closure $\bar{A} \subset T$ is a closed subgroup of T and must thereby be a closed submanifold, by the closed subgroup lemma for Lie groups. Dimension 0 is ruled out since by Kronecker's theorem the first coordinate nr projects to a dense set in \mathbb{R}/\mathbb{Z} . If \bar{A} has dimension 2, then A clearly has points in the triangle Δ defined by $1 - \langle nr \rangle \geq \langle ns \rangle \geq \langle nr \rangle$. If \bar{A} has dimension 1 then it is a finite union of lines of rational slope, including one through $(0, 0)$. The slope can be neither 0 nor ∞ , since neither s nor r is an integer. It is then elementary to see that A must intersect Δ .

Editorial comment. The other two solvers effectively gave self-contained *ad hoc* proofs of the above claim, using Kronecker lemma arguments. One would-be solver showed that for every rational r there is an interval $I = [r, r + E_r)$, such

that for any $r_0 \in I$, $\varphi(n, r_0)$ is not always > 1 , but then jumped to the fallacious conclusion that these intervals must cover all of \mathbb{R} , owing to the density of the rationals.

Solved also by R. J. Chapman (U.K.) and the proposer. Two incorrect solutions were received.

Nonnegative Curve Fitting

10219 [1992, 362]. *Proposed by Alan Horwitz, Penn State University, Media PA.*

(a) Suppose that the function f is positive on \mathbb{R} and that $f''(x)$ exists for all $x \in \mathbb{R}$. Prove that there exists $x_0 \in \mathbb{R}$ such that the second order Taylor polynomial of f centered at x_0 is also positive on \mathbb{R} .

(b)* Let n be an arbitrary even integer, and suppose that f is positive on \mathbb{R} and that $f^{(n)}(x)$ exists for all $x \in \mathbb{R}$. Does there exist $x_0 \in \mathbb{R}$ such that the n -th order Taylor polynomial of f centered at x_0 is also positive on \mathbb{R} ?

Solution of (a) by David Callan, University of Wisconsin, Madison, WI. By completing the square, it is easy to verify that x_0 has the desired property iff $f'(x_0) = f''(x_0) = 0$ or $f(x_0)f''(x_0) > f'(x_0)^2/2$. Let $g(x) = \sqrt{f(x)}$, thus $g'(x) = f'(x)/(2g(x))$ and $g''(x) = (f(x)f''(x) - f'(x)^2/2)/(2g(x)^3)$. If there exists x_0 with $g''(x_0) > 0$, clearly this x_0 will do. If not, $g''(x) \leq 0$ for all x and the graph of g lies entirely on or below each of its tangent lines. Hence since g is positive everywhere, these tangents must all be horizontal, so g and f are constants, and any x_0 will do.

Editorial comment. All solvers confined attention to part (a). In the original proposal, the proposer included a reference to J. Briggs & L. Rubel, "Interpolation by non-negative polynomials", *J. Approx. Theory* 30 (1980), 160–168 for a version of part (b) on a compact interval. He also mentioned A. Horwitz, "Interpolation by polynomials non-negative on the real line", *preprint* as treating the analogous problem in which Taylor polynomials are replaced by interpolating polynomials at distinct points.

Solved also by K. F. Andersen (Canada), D. Caccia, R. J. Chapman (U.K.), R. Cooke, E. A. Herman, O. P. Lossers (The Netherlands), M. Mócsy (Hungary), A. Nijenhuis, K. Schilling, Anchorage Math Solutions Group, and the proposer.

A Little Measure Theory

10225 [1992, 462]. *Proposed by Paul R. Chernoff, University of California, Berkeley, CA.*

Suppose that $\phi: [0, \infty) \rightarrow [0, \infty)$ is a strictly increasing, strictly concave function with $\phi(0) = 0$. Let m^* be Lebesgue outer measure on the unit interval $I = [0, 1]$. For $E \subset I$, define $n^*(E) = \phi(m^*(E))$. Show that n^* is an outer measure and determine the n^* -measurable sets.

Solution by Michael B. Gregory, University of North Dakota, Grand Forks, ND. The monotonicity of n^* follows from that of m^* and the fact that ϕ is increasing; $n^*(\emptyset) = 0$ is a consequence of $m^*(\emptyset) = 0$ and $\phi(0) = 0$. To verify that n^* is countably subadditive, we will first show $\phi(t)/t$ is strictly decreasing on $(0, \infty)$: let

$0 < a < b$ be given. Because ϕ is strictly concave, with $\alpha = a/b$, we have

$$\phi(a) = \phi((1 - \alpha)0 + ab) > (1 - \alpha)\phi(0) + \alpha\phi(b) = \frac{a}{b}\phi(b).$$

Thus, $\phi(a)/a > \phi(b)/b$, as claimed. In turn, this implies that ϕ is *strictly subadditive*: if $a, b > 0$ then $\phi(a + b) < \phi(a) + \phi(b)$. To see this, write

$$\begin{aligned}\phi(a + b) &= \frac{a}{a + b}\phi(a + b) + \frac{b}{a + b}\phi(a + b) \\ &< a\frac{\phi(a)}{a} + b\frac{\phi(b)}{b} = \phi(a) + \phi(b).\end{aligned}$$

Now one can see, by induction, that if $x_1, x_2, \dots, x_k > 0$ then $\phi(x_1 + \dots + x_k) < \phi(x_1) + \dots + \phi(x_k)$. It follows that if $\langle x_k \rangle$ is any sequence of nonnegative numbers such that the sequence $\langle \phi(x_k) \rangle$ is any sequence of nonnegative numbers such that the sequence $\langle \phi(x_k) \rangle$ is summable then the sequence $\langle x_k \rangle$ is also summable and

$$\phi\left(\sum_{k=1}^{\infty} x_k\right) \leq \sum_{k=1}^{\infty} \phi(x_k),$$

because ϕ is continuous at each positive number. (Concave functions are continuous at the interior points of their intervals of concavity.) Therefore, since ϕ is increasing, if $\langle E_k \rangle$ is any sequence of subsets of I , then (we may assume that the sequence $\langle \phi(m^*(E_k)) \rangle$ is summable)

$$n^*\left(\bigcup_{k=1}^{\infty} E_k\right) \leq \phi\left(\sum_{k=1}^{\infty} m^*(E_k)\right) \leq \sum_{k=1}^{\infty} \phi(m^*(E_k)) = \sum_{k=1}^{\infty} n^*(E_k),$$

and consequently n^* is an outer measure on I .

We now show that the n^* -measurable subsets of I are exactly the m^* -measurable sets $E \subseteq I$ with $m^*(E) = 0$ or $m^*(E) = 1$. Because the n^* -measure of such a set or its complement (in I) is zero, it follows from the basic theory of outer measures, that every such set is n^* -measurable.

To prove the converse, we first verify, by contrapositive, that any n^* -measurable set must be m^* -measurable. Suppose that $E \subseteq I$ is *not* m^* -measurable. Then there is a *test set* $A \subseteq I$ such that (with E^c denoting the complement of E in I)

$$m^*(A) < m^*(A \cap E) + m^*(A \cap E^c).$$

Because ϕ is strictly increasing, it follows that

$$\begin{aligned}\phi(m^*(A)) &< \phi(m^*(A \cap E) + m^*(A \cap E^c)) \\ &< \phi(m^*(A \cap E)) + \phi(m^*(A \cap E^c));\end{aligned}$$

the last inequality follows from the strict subadditivity of ϕ . This means that

$$n^*(A) < n^*(A \cap E) + n^*(A \cap E^c)$$

and hence E is not n^* -measurable. It follows that all the n^* -measurable sets must be Lebesgue measurable.

Now assume $E \subseteq I$ is a Lebesgue measurable set such that $0 < m^*(E) < 1$. Taking the “test set” $A = I$, we see that because both $m^*(A \cap E)$ and $m^*(A \cap$

E^c) are positive,

$$\begin{aligned} n^*(A) &= \phi(m^*(A)) = \phi(m^*(A \cap E) + m^*(A \cap E^c)) \\ &< \phi(m^*(A \cap E)) + \phi(m^*(A \cap E^c)) \\ &= n^*(A \cap E) + n^*(A \cap E^c). \end{aligned}$$

Therefore, E is not n^* -measurable. Thus, a subset $E \subseteq I$ is n^* -measurable if and only if either $m^*(E) = 0$ or $m^*(E) = 1$. This means that the family of n^* -measurable sets is the σ -algebra in I generated by the subsets (of I) that have Lebesgue measure zero.

Solved also by R. J. Chapman (U.K.), N. Komanda, and the proposer. Two incomplete solutions were also received.

Characterizing The Ball

10228 [1992, 463]. *Proposed by Ernesto Bruno Cossi and Marcos Antonia Sebastiani, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.*

(a) Let \mathcal{X} be a Banach space, and let B be a bounded, nonempty subset of \mathcal{X} such that, for any pair of points x and y in B , there is an open ball U such that $U \subset B$, $x \in U$ and $y \in U$. Show that B is an open ball.

(b) Show that the result of part (a) does not generalize to the case in which \mathcal{X} is only assumed to be a complete metric space.

Solution by Robert B. Israel, University of British Columbia, Vancouver, B.C., Canada. Part (a): Let $R = \sup\{r: B \text{ contains an open ball of radius } r\}$. This is finite since B is bounded. Let $\langle U_n \rangle$ be a sequence of open balls contained in B , with radii $r_n \rightarrow R$ as $n \rightarrow \infty$. Let x_n be the centre of U_n . Note that

$$\text{diam}(U_i \cup U_j) = r_i + r_j + \|x_i - x_j\|. \quad (1)$$

Thus there exist open balls contained in B (containing pairs of points in $U_i \cup U_j$) with diameters arbitrarily close to $r_i + r_j + \|x_i - x_j\|$. Since the diameter of a ball is twice its radius, we must have $r_i + r_j + \|x_i - x_j\| \leq 2R$, i.e., $\|x_i - x_j\| \leq 2R - r_i - r_j$. Therefore $\langle x_n \rangle$ is a Cauchy sequence, converging to some point x . I claim that B is the open ball V of radius R centered at x .

Now for any $r < R$, the open ball of radius r centered at x is contained in U_n , and therefore in B , if $\|x_n - x\| + r < r_n$, and this is true for n sufficiently large. Since V is the union of these balls, $V \subseteq B$. On the other hand, if $\|z - x\| > R$ there is a point $u \in V$ with $\|u - z\| > 2R$. Thus $u \in B$, but any open ball containing u and z must have radius greater than R , so $z \notin B$. This means that B is contained in the closure of V . But clearly B is open, so $B = V$.

Part (b): Consider the complete metric space X consisting of five points a, b, c, d and e with $d(a, b) = d(b, c) = d(c, d) = d(a, d) = 2$, $d(a, c) = d(b, d) = 4$, and $d(x, e) = 3$ for all $x \neq e$. Let $B = \{a, b, c, d\}$. Note that an open ball contained in B consists of either one point or any three points of B . Thus B has the required property, but is not an open ball.

Editorial comment. G. Muraz & P. Szeptycki remarked that the solution to (a) remains valid in any complete metric space satisfying (1). Nasha Komanda, Reiner Martin, and Kenneth Schilling also provided solutions to (b) in which \mathcal{X} has five points and B has four points. Another popular collection of examples used subsets

of the plane. G. Muraz & P. Szeptycki took $\mathcal{X} = \{(s, 0): -1 \leq s \leq 1\} \cup \{(0, 1)\}$ and $B = \{(s, 0): -1 < s \leq 1\}$. The example given by Jean-Pierre Grivaux and Dave Trautman took $\mathcal{X} = \mathbb{R}$ with the metric given by $d(x, y) = |x - y|/(1 + |x - y|)$, with B being the set of positive real numbers.

Solved also by J. Alexopoulos, S.-K. Chung, M. Dindos (Slovakia), J.-P. Grivaux (France), R. Holzsgager, E. G. Katsoulis, K. S. Kedlaya (student), N. Komanda, R. Martin (student), M. Mócsy (Hungary), G. Muraz & P. Szeptycki (France), K. Schilling, D. Trautman, and the University of Wyoming Problem Circle. A solution by Marcos Antonio Sebastiani accompanied the original proposal.

Some Inverse Hilbert Polynomials

10240 [1992, 674]. *Proposed by Michael Golomb, Purdue University, West Lafayette, IN.*

Fix an integer n . For each integer m with $0 \leq m \leq n$, let p_m be a polynomial of degree n for which $\int_0^1 p_m(x) x^l dx = 0$ for $0 \leq l \leq n$ with $l \neq m$, while $\int_0^1 p_m(x) x^m dx = 1$.

(a) Determine the value of $\int_0^1 p_m^2(x) dx$.

(b) Find an explicit expression for p_m and prove that the coefficient of x^l in p_m is the same as the coefficient of x^m in p_l for $0 \leq l < m \leq n$.

Solution by Robin J. Chapman, University of Exeter, Exeter, U.K.. Let

$$c_m = \frac{(-1)^m (n+m+1)!}{m!^2 (n-m)!} \text{ and } c_{l,m} = \frac{c_l c_m}{l+m+1}.$$

(a) We have

$$\int_0^1 p_m^2(x) dx = c_{m,m} = \frac{(n+m+1)!^2}{m!^4 (n-m)!^2 (2m+1)}.$$

(b) Also

$$p_m(x) = \sum_{l=0}^m c_{l,m} x^l.$$

Proof: Represent a polynomial f of degree $\leq n$ by a row vector $v_f = (a_0, a_1, \dots, a_n)$, where $f = \sum_{m=0}^n a_m x^m$. Then $\int_0^1 f(x)g(x) dx = v_f H v_g^T$ where $H = (1/(l+m+1))_{l,m=0}^n$ is the $n+1$ by $n+1$ Hilbert matrix. If $C = (c_{l,m})_{l,m=0}^n = H^{-1}$ then putting $p_m(x) = \sum_{l=0}^n c_{l,m} x^l$ we immediately get $\int_0^1 p_m(x) x^l dx = \delta_{l,m}$. As H is symmetric so is C , and so the latter assertion of (b) is immediate. Also

$$\int_0^1 p_m(x)^2 dx = \sum_{l=0}^n c_{l,m} \int_0^1 p_m(x) x^l dx = c_{m,m}.$$

The formula for $c_{l,m}$ is classical.

Editorial comment. A reference for the entries of H^{-1} supplied by readers was D. E. Knuth, *The Art of Computer Programming*, Vol. I, Addison-Wesley, 1973, ex. 41, p. 36, which deals with the more general *Cauchy matrix*, and ex. 45, p. 37. David and Peter Borwein observed that the result may be generalized to *Müntz polynomials* $\sum_{k=0}^n a_k x^{\lambda_k}$ where the exponents λ_k are real numbers satisfying

$-1/2 < \lambda_0 < \lambda_1 \cdots < \lambda_n$. The quantities

$$c_m = \frac{\prod_k \{\lambda_m + \lambda_k + 1: 0 \leq k \leq n\}}{\prod_k \{\lambda_m - \lambda_k: 0 \leq k \leq n; k \neq m\}}$$

generalize the corresponding expressions in the classical case, and the coefficient $c_{l,m} = c_l c_m / (\lambda_l + \lambda_m + 1)$. The expression for the inner products of Müntz polynomials in terms of their coefficients is given by a Cauchy matrix, so the approach used above applies in this generality. Another approach to part (a), suggested by the proposer, is to calculate the coefficients of the $p_m(x)$ in terms of the orthogonal basis of Legendre polynomials.

Solved also by S. Ali, R. Bagby, S. F. Barger, D. J. Barrett, K. L. Bernstein, D. Borwein & P. Borwein (Canada), D. Callan, R. Cerf (France), D. A. Darling, K. Diethelm (Germany), I. Dimitrić, P. G. Engstrom, J. Fukuta (Japan), J. A. Gomez Ortega (Mexico), E. A. Herman, R. Holzinger, W. Hong, F.-A. Izadi (Iran), I. Kastanas, N. Komanda, H. K. Krishnapriyan, C. Libis, O. P. Lossers (The Netherlands), S. Matz, A. Nijenhuis, A. Pedersen (Denmark), F. C. Rembis, H. J. Seiffert (Germany), D. Spellman, M. Stamp, W. F. Trench, D. B. Tyler, R. L. Van de Wetering, M. Vowe (Switzerland), National Security Agency Problems Group, Skidmore College Problem Group, Western Maryland College Problems group, University of Wyoming Problem Circle, and the proposer.

Collaborating editors: *David F. Appleyard, Paul T. Bateman, Duane M. Broline, Barry W. Brunson, Frank S. Cater, Gulbank D. Chakerian, Underwood Dudley, Gerald A. Edgar, Michael A. Filaseta, Ira M. Gessel, Richard A. Gibbs, Jerrold R. Griggs, Douglas A. Hensley, John R. Isbell, Mourad E. H. Ismail, Murray Klamkin, Daniel J. Kleitman, Frederick W. Luttman, Frank B. Miles, Richard Pfiefer, Stephen L. Portnoy, J. O. Shallit, John Henry Steelman, Kenneth B. Stolarsky, David E. Tepper, Douglas B. Tyler, Daniel Ullman, and William E. Watkins.*

From the *New Yorker*, July 26, p. 62:

This was a world bounded by a diminishing set of coordinates. There were from the beginning a finite number of employers who needed what these people knew how to deliver, and what these people knew how to deliver was only one kind of product. "Our industry's record at defense conversion is unblemished by success" . . .

Contributed by Emma Lehmer
1180 Miller Avenue
Berkeley, CA 94708

REVIEWS

Edited by **Darrell Haile**
Indiana University, Bloomington, IN 47405

Reality Rules I. The Fundamentals; II The Frontier. By John L. Casti, John Wiley & Sons, New York, 1992, \$72.00 (set).

Reviewed by **Rutherford Aris**

Par ma foi! il y a plus de quarante ans que je dis de la prose sans que j'en susse rien.

Many a mathematical practitioner must have reacted—albeit with a whimsical smile—to the current popularity of “mathematical modeling” as M. Jourdain does when told that his orders to Nicole are indeed prose. Books with these words in their titles or subtitles have been sprouting plentifully in the last decade and more. They range from elementary introductions and collections of case histories through general discussions of modeling methodology to the metaphysics and philosophy of the subject. Such topics as chaos, fractals and complexity have been treated at book-length at a level almost accessible to the notorious ‘man in the street’. With this popularization, often very well done, there has come a healthy awareness of the more concrete aspects of mathematics, even though fraught with the danger of misunderstanding which either invests the model with a spurious authority or else dismisses it as something less than real. (Some recent observations by Davis and Boose-Bavenbek in *SIAM News*, 6 May 1993, pp. 6–7, on ‘Mathematics and the Media’ are very much to the point.) Practitioners of the art and craft of modeling (I use ‘and,’ for the old disdain of the artist for the craftsman is as sterile as that of the ‘pure’ mathematician for the applied which sometimes used to fester among the second-rate. I do not regard Hardy’s famous *Apology* as expressing disdain, even though there is clearly a lack of sympathy.) are often too much wrapped up in the detail and the peculiar beauties of their model to be easily able to step aside and take a larger view. There is therefore a genuine need for expositions of the general principles and overarching considerations of modelling that will help us to get a sound perspective on the whole enterprise. It is to this need that Casti’s two volume extension of his 1989 *Alternate Realities* is addressed.

Reality Rules can be used independently of its precursor and it is very much more than a mere revision or updating of it. Its title is deliberately ambivalent, its scope far-reaching and its style somewhat reminiscent of the late Dick Bellman to whose memory it is dedicated. I mean this in the best sense, not that it is imitative. But there is the same variousness of example, introduced seriously and thoughtfully enough, but leaving much unsaid and needing to be supplemented from the references, given with a useful word or two of comment at the chapter’s end. Mention of some of the examples will indicate the range of contacts between the external reality and mathematical model: Keynesian dynamics and equivalent economies, forest insect pest control, the van der Pol equation, racial integration

in urban housing, rodents and cellular automata, DNA sequences, chaos in monetary aggregates, the dynamics of beer distribution, the art of Escher and Mondrian, industrial manufacturing processes, the discrete, controlled pendulum, cognitive theory, sea bass spawning and the flowering times of competing plants. It is an abundantly furnished table to which we are led. A very strong point of the book is the generous supply of exercises, discussion questions and problems which make the book attractive for classroom or seminar use. There is to be a solutions manual, obtainable from the publishers.

Because the individual models in *Reality Rules* are introduced to illuminate some particular point or other, only a few are developed far enough that one gets a sense of the shaped and polished model. Of course there is a certain danger in even wanting a model to be a complete thing. No model stands alone and there are always plenty of possible developments. But provided the sin of idolatry is avoided, a good model has a life of its own and may be enjoyed for its own sake. Take, for instance, the Lorenz equations, the logistic map or some of Otto Rössler's systems. They have all the 'juice and joy' of old, and modern, masters and one may well seek, with Hopkins, to know whence it comes. But perhaps this must await a mathematical editor of the temperament of a classical scholar who will bring out a *Corpus Systemorum Dynamicorum* to rival the greatness of the *Corpus Vasorum* or Lowe's *Codices Latinae Antiquiores*.

Though the details of encoding—i.e. the task of setting up a model with its choice of variables and parameters, their manipulation and the preliminary shaping of the equations—are not the subject of *Reality Rules*, the modeling relation is clearly explained and the system scientist installed as the ferryman across the river that divides the mathematical and the 'real' worlds, encoding the natural system in one direction and decoding the formal system in the other. Three major types of model are considered in the "Fundamentals" volume and they are related to major classes of problems. Thus the singularities of resource systems are amenable to analysis from the viewpoint of catastrophe theory; pattern formation and the emergence of living forms link with discrete dynamical systems and cellular automata; turbulence and population dynamics invite chaos theory. Similarly in the second volume, "Frontiers", —the dividing line between the volumes is clearly not to be taken as hard and fast—games and the theory of evolution occupy chapter five, while chapter six is a system-theoretic view of minds and mechanisms and seven deals with control, optimal, adaptive and anticipatory. Not surprisingly, almost all the examples up to this point have been dynamical systems, but in chapter eight he introduces the geometrical concept of a simplicial complex as a tool in teasing out the inner structure of a natural system. The homology of such complexes leads to a natural definition of connectivity and this is illustrated by an analysis of the Middle East situation, Escher's *Sky and Water* and a work of Piet Mondrian. There is a nice tie back into dynamical systems in a discussion of the structure of reachability and a range of discussion and other problems touching on every thing from *The Death of a Salesman* to elementary homology theory and from the Betti numbers of the Middle East conflict to the food web of *Nepenthes albonmarginata*. This chapter illustrates the strength and weakness of the work. One marvels at and is stimulated by Casti's Bellmanesque fecundity, but the exposition of the underlying theory is necessarily so brief that even the copious references leave one wondering just where to start digging in the mathematical soil that holds promise of so interesting a harvest.

In the final chapters of this long and challenging work, Casti tackles the truly philosophical question of how we know. First he raises the question of whether

there are natural limits to reason, by looking at the properties of the Turing machine and the import of Gödel's Incompleteness Theorem and finishes with a chapter on the role of myths, models and paradigms. These are deep waters where few of us would claim to have any final insight, and where indeed such a claim would be viewed with suspicion. Casti does a good job of laying out several of the recent approaches and ends with the wise remark that there are no complete answers adding the admonition that "the only rule in the Reality Game is to avoid falling into that most common of all human delusions, the delusion of a single reality—our own!"

In summary these volumes provide a timely and insightful overview of mathematical modeling, rich in example and lively in presentation. It should go far to ensure that the growing self-consciousness of the mathematical modeler is the eager self-edification of the honest artisan, rather than the empty self-congratulation of the *bourgeois gentilhomme*.

Chemical Engineering and Mathematical Sciences
University of Minnesota
Minneapolis, MN 55455-0373

Geometry of Surfaces. By John Stillwell. Springer-Verlag, New York, 1992.

Reviewed by **David L. Webb**

It is pretty generally recognized that modern Riemannian geometry is a subject which has come of age during this century, following the development of foundational ideas able to bear the weight of the intuitive geometric arguments pioneered by Riemann and others. It is also a subject whose influence is felt nearly everywhere, not merely in mathematics but also in physics. Many important physical theories are most naturally expressed in geometric terms: general relativity relies upon Lorentzian geometry, classical mechanics is naturally formulated in terms of symplectic geometry, and modern gauge field theories make heavy use of the theory of connections on fibre bundles. The interactions with other parts of mathematics are equally striking: one well-known example is the classical Uniformization Theorem for Riemann surfaces and its 3-dimensional counterpart, Thurston's visionary Geometrization program for understanding the topology of 3-manifolds.

However, despite its pivotal role in many areas of mathematics and physics, modern geometry is still not a standard part of the undergraduate curriculum at most institutions. Indeed, geometry tends to fare rather poorly at the undergraduate level: in many colleges and universities, if there is a regular geometry course, it is likely to be one intended for future secondary school teachers - a course chiefly concerned with questions of axiomatics, which treats non-Euclidean geometry from a "synthetic" viewpoint and perhaps introduces the Poincaré plane, but does little to lay the groundwork for an understanding of the spectacular successes of modern Riemannian geometry. This is unfortunate, as there are a number of really excellent texts now available. Three particularly nice examples are *Differential Geometry of Curves and Surfaces* by Manfredo do Carmo (Prentice-Hall, 1976),

Elementary Differential Geometry by Barrett O'Neill (Academic Press, 1966), and *Riemannian Geometry-A Beginner's Guide* by Frank Morgan (Jones and Bartlett, 1992).

These observations raise two natural questions: Why does geometry tend to get such short shrift in the traditional undergraduate curriculum? And why is a new book needed?

There are many answers to the first question. One obvious answer is that there is an intrinsic obstacle to doing geometry which always seems to rear its head: although the ideas are often clear, beautiful, and compelling, making them precise (and avoiding the mistakes which can easily arise from a free-wheeling use of "geometric intuition") requires some machinery which seems far too sophisticated given the simplicity of the underlying ideas; this arises chiefly because one must introduce coordinates and hence must keep track of what happens when one changes the coordinates. (Anyone who has taught linear algebra frequently is aware of the difficulties students have in mastering the yoga of basis change and similarity, even in the rather temperate climate of a linear space; in the more forbidding climes of a nonlinear object such as a manifold, the pedagogical difficulties are that much worse.) A related but less obvious difficulty arises from the nature of the underlying objects: one is usually interested in *intrinsic geometry*, the study of spaces whose geometry arises from an intrinsic notion of distance rather than being "inherited" from the geometry of an ambient Euclidean space. In undertaking such a study, it is of course natural to try to set up the foundations in such a way that one can do calculus in a nonlinear space without relying upon an ambient linear space...but this takes some work: it leads one to introduce a panoply of objects (including abstract manifolds and vector bundles) which tend to bewilder the neophyte and to lend credence to the jocular pronouncement that differential geometry is the study of those properties which are invariant under change of notation. An alternative course is to begin with manifolds embedded in Euclidean space, then gradually phase the ambient space out of the picture (thereby, as it were, removing the Cheshire Cat and leaving only the grin remaining). This approach is perhaps sounder pedagogically (and is the one adopted in the three books cited above); however, it tends to distract one from the intrinsic character of the study.

Stillwell's book circumvents these difficulties by restricting its focus to a special case, but one which is quite interesting and leads to much beautiful classical geometry: the study of surfaces of *constant curvature*. (This is much easier, just as it is much easier to define what is meant by "Lebesgue measure zero" than to define what is meant by "Lebesgue measure".) The point of view is that of Klein's program of studying geometry via the study of groups of symmetries. The basic objects of study, rather than being surfaces lying in some Euclidean space from which they derive their geometry, are now *quotients* of the Euclidean plane, of the hyperbolic plane, or of the round 2-sphere by discrete groups of isometries acting without fixed points. In this approach, one is doing "intrinsic" geometry right at the outset, since these are naturally quotients rather than subobjects of familiar geometric spaces. Restricting to the case of constant curvature relieves one of the necessity of rounding up the usual suspects (vector fields, forms, frame fields, shape operators, connections, etc.) before embarking; one can easily formulate basic notions such as "geodesic" and "area" locally by lifting them to the familiar geometry of the covering surface. The main role is thus played not by the machinery of calculus, but by that of group theory (note that the word "differential" does not appear in the title).

The book features leisurely and informative treatments of the three kinds of geometry considered: Euclidean (curvature $\equiv 0$), spherical (curvature $\equiv 1$), and hyperbolic (curvature $\equiv -1$). In each case, one begins with a study of the isometries of the simply connected space having the desired constant curvature and then proceeds with the study of general surfaces by means of a study of the discrete subgroups of the group of isometries. Of the three, hyperbolic geometry is perhaps the most interesting, since it is the geometry of “most” Riemann surfaces, and the connections with complex analysis are especially appealing; appropriately enough, it gets the lion’s share of attention. The sections labeled “Discussion” are especially noteworthy: they furnish brief heuristic introductions to some fascinating connections with other areas, including elliptic functions, modular forms, and Riemann surfaces. Later on, the book includes some basic material about the topological classification of surfaces and the fundamental group. When pursuing the “Riemannian covering” approach to geometry adopted here, it is quite natural to inquire what happens if one discards the requirement that the groups of isometries act without fixed points. This leads to the idea of a Riemannian *orbifold* of constant curvature, a space which locally looks geometrically like the quotient of a simply connected space of constant curvature by a finite group of isometries; this idea is introduced in the final chapter. It should also be noted that the book reads quite smoothly and is liberally furnished with helpful pictures.

This summary suggests an answer to the second question: Why is a new book needed? One answer is that the approach taken here is significantly different from that of most available texts written at a comparable level, and it furnishes comparatively easy access to some very beautiful mathematics. One of the frustrating things about teaching a one-quarter differential geometry course from one of the standard texts is that one scarcely has time to introduce the basic machinery before the term is over, and hence one never really gets very far into the *geometric* heart of the subject. For such a course, this book fills the bill admirably. Of course, there are inevitable drawbacks as well: the book neatly sidesteps the difficulties of teaching differential geometry by the rather brutal expedient of excising the “differential” to concentrate on the “geometry”, so that a student wishing to go further (say, to study traditional Riemannian geometry or general relativity) will be ill-equipped to do so; such a student would be well-advised to peruse do Carmo, O’Neill, Morgan, or even one of the many more advanced treatments of the subject to gain familiarity with the requisite machinery. Thus Stillwell’s book should be viewed as complementary to (rather than as a substitute for) the standard treatments, and in this role it is a welcome contribution. Most important, however, this book is tremendously valuable as a reminder of *why* geometry is so captivating in the first place.

Department of Mathematics
Dartmouth College
Hanover, NH 03755

Answer to Picture Puzzle
(p. 139)
Alfred Haar.

TELEGRAPHIC REVIEWS

Edited by Arnold Ostebee and Paul Zorn

with the assistance of the Mathematics Departments of
Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T : Textbook	P : Professional Reading	1-4: Semester
C : Computer Software	L : Undergraduate Library	** : Special Emphasis
S : Supplementary Reading	13: Grade Level	?? : Questionable

Readers are advised that price information is subject to change. Selected books and software packages receive a second, more extensive review in the *Monthly*.

Books and software submitted for review should be sent to *Book Reviews Editor*, *American Mathematical Monthly*, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098.

General, T(15-16: 1), S, P, L. *Laws of the Game: How the Principles of Nature Govern Chance.* Manfred Eigen, Ruthild Winkler. Transl: Robert & Rita Kimber. Princeton Univ Pr, 1993, xv + 347 pp, \$16.95 (P). [ISBN 0-691-02566-5] From the Foreword: "Everything that happens in our world resembles a vast game in which nothing is determined in advance but the rules, and only the rules are open to objective understanding." BC

Logic, P. *Gödel's Theorems: A Workbook on Formalization.* Verena Huber-Dyson. Teubner-Texte zur Mathematik, B. 122. BG Teubner Stuttgart, 1991, 292 pp, (P). [ISBN 3-8154-2023-7] Detailed explication of Gödel's Theorems for mathematicians and philosophers. Includes introduction to intuitionistic reasoning and topos theory, examples from combinatorial group theory. Exercises and essay suggestions. KES

Logic, T(13-14: 1). *Clear Thinking: An Invitation to Logic.* Gary Jason. Jones & Bartlett, 1992, x + 518 pp, \$30 (P). [ISBN 0-86720-182-7] Text for introductory logic course (not mathematical logic). KES

Foundations, T*(14: 1), S, L. *Sets, Functions and Logic: An Introduction to Abstract Mathematics, Second Edition.* Keith Devlin. Chapman & Hall, 1992, xi + 147 pp, \$22.50 (P). [ISBN 0-412-45980-9] Readable, concise introduction to logic, sets, functions, relations, complex numbers. Good exercises; no solutions in book. (First Edition, TR, March 1982.) KES

Combinatorics, T(15-16: 2), L. *Aspects of*

Combinatorics: A Wide-Ranging Introduction. Victor Bryant. Cambridge Univ Pr, 1993, viii + 266 pp, \$64.95. [ISBN 0-521-41974-3] Emphasizes graph and transversal theory. Chapters treat Latin squares, the marriage theorem, rook polynomials, planar graphs, Ramsey theory; generating functions don't appear. Exercises, hints, and solutions. LC

Discrete Mathematics, T*(13-14: 1). *Discrete Mathematics, Second Edition.* John A. Dossey, et al. HarperCollins College, 1993, xv + 555 pp, \$40. [ISBN 0-673-46287-0] Revised chapters on graphs, trees, recurrence; new material on generating functions; supplementary exercises and computer projects for each chapter. (First Edition, TR, August-September 1987; Extended Review, February 1988.) KES

Discrete Mathematics, T(13-14: 1), L. *Discrete Mathematics for Computing.* John E. Munro. Chapman & Hall, 1992, x + 306 pp, (P). [ISBN 0-412-45650-8] Concise text covers integers, number representation, sets, functions, relations, logic, proof, recursion, analysis and correctness of algorithms, graphs and trees, counting, algebraic structures. Introduction to each chapter explains relevance of topic to computer science. Algorithms in English. KES

Number Theory, T*(16-17: 2), L. *Introduction to Elliptic Curves and Modular Forms, Second Edition.* Neal Koblitz. Grad. Texts in Math., V. 97. Springer-Verlag, 1993, x + 248 pp, \$49. [ISBN 0-387-97966-2] New edition of an exceptional introduction to elliptic curves. Changes focus on new developments concerning the Birch and Swinnerton-

Dyer conjectures. (*First Edition*, TR, April 1985.) Highly recommended. MPR

Linear Algebra, T*(14: 1), L. *Elementary Linear Algebra*, Second Edition. Roland E. Larson, Bruce H. Edwards. DC Heath, 1991, xiv + 670 pp, \$36 net. [ISBN 0-669-24592-5] Excellent applications, many exercises (some theoretical), historical notes. New chapters on complex spaces, linear programming. (*First Edition*, TR, December 1988.) CEC

Algebra, T(17–18: 1, 2), L. *Elementary Abstract Algebra*. Lawrence E. Spence, Charles Vanden Eynden. HarperCollins College, 1993, xviii + 456 pp, \$42. [ISBN 0-673-38583-3] Starts with ring theory (elementary properties, subrings, direct sums, homomorphisms and isomorphisms), then fields, integral domains, extension fields; all before group theory. Worth a look. LC

Calculus, T(13), L. *Calculus*. Deborah Hughes-Hallett, Andrew M. Gleason. Wiley, 1994, xviii + 685 pp, (P). [ISBN 0-471-31055-7] The “Harvard Consortium materials.” Now in attractive soft-cover form with the customary inexplicable photograph on the cover (a sky-diver, in this case). Only slightly less lean than in its preliminary edition (partial fractions and a glancing contact with the Mean Value Theorem have been added). Not isomorphic to the usual calculus book: no epsilons and deltas, few proofs. Many standard topics (e.g., trig substitution) are missing, but the problems and the text give students much more support than usual in attaining an intuitive understanding of the ideas of calculus. Get the book, and think about the trade-offs yourself. JO

Calculus, T(13–14: 3), C, L. *Calculus Using Mathematica*. K.D. Stroyan. Academic Pr, 1993, xxv + 532 pp, \$50 with disk, [ISBN 0-12-672972-7]; *Scientific Projects and Mathematical Background*, xi + 353 pp, (P). [ISBN 0-12-672975-1] In four parts: core text, science projects, mathematical projects, and Mathematica NoteBooks. Introduces calculus ideas through intriguing economics, science, and engineering applications. Use of Mathematica and student projects are essential for any course based on the text. Very interesting, albeit somewhat nontraditional. AO

Complex Analysis, T(17–18: 1, 2), P*. *Normal Families*. Joel L. Schiff. Universitext. Springer-Verlag, 1993, xii + 236 pp, \$39 (P). [ISBN 0-387-97967-0] First book-length study of normal families of analytic functions since introduction in 1920’s by Paul Montel. Relatively self-contained: covers basics, analytic theory, meromorphic theory, Bloch

principle, current applications. Scope ranges from early graduate through research levels. Many references; no exercises. PZ

Differential Equations, S(14–15), C*. *MacMath 9.2: A Dynamical Systems Software Package for the Macintosh*. John H. Hubbard, Beverly H. West. Springer-Verlag, 1993, vii + 162 pp, \$49.95 (P), with disk. [ISBN 0-387-94135-5] Software updated for System 7 operating system. (MacMath 9.0 owners may update to Version 9.2 by contacting authors). AO

Differential Equations, P. *Nonlinear Stokes Phenomena*. Ed. Yu. S. Il’yashenko. Adv. in Soviet Math., V. 14. AMS, 1993, xiv + 287 pp, \$116. [ISBN 0-8218-4112-2] 5 papers survey recent developments.

Dynamical Systems, P. *Encounter with Chaos: Self-Organized Hierarchical Complexity in Semiconductor Experiments*. J. Peinke, et al. Springer-Verlag, 1992, x + 289 pp, \$59. [ISBN 0-387-55647-8] Nonlinear behavior of low-temperature impact ionization breakdown in intrinsic germanium serves as experimental background for the development of chaotic dynamics. Good for experimenters who want to learn more about mathematics of nonlinear dynamics, and for mathematicians who want to learn about applications of chaos theory. SP

Analysis, T(18: 1, 2), P. *Difference Equations and Inequalities: Theory, Methods, and Applications*. Ravi P. Agarwal. Pure & Appl. Math., V. 155. Marcel Dekker, 1992, xiii + 777 pp, \$150. [ISBN 0-8247-8676-9] Comprehensive treatment develops discrete versions of Rolle’s, mean value, Kneser’s theorems; Taylor’s formula; l’Hôpital’s rule. SP

Analysis, T*(17: 2). *Real and Functional Analysis, Third Edition*. Serge Lang. Grad. Texts in Math., V. 142. Springer-Verlag, 1993, xiv + 580 pp, \$49.50. [ISBN 0-387-94001-4] Successful text has been reorganized; integration now precedes functional analysis. Begins with topics from point set topology, covers the usual topics in a solid real analysis course followed by a rigorous development of key ideas in functional analysis. Well-conceived, clearly presented. Nice problem sets throughout. (1983 edition published as *Real Analysis*, TR, November 1983.) TAV

Algebraic Geometry, T(16–18: 1, 2), S, P, L. *Gröbner Bases: A Computational Approach to Commutative Algebra*. Thomas Becker, Volker Weispfenning, Heinz Kredel. Grad. Texts in Math., V. 141. Springer-Verlag, 1993, xxii + 574 pp, \$49. [ISBN 0-387-97971-9] The theory of Gröbner bases is an analogue, for multivariate polynomials, of the Euclidean algorithm

for computing gcd's. Advances in symbolic computation have rendered eminently practical what used to be primarily of theoretic interest. BC

Differential Geometry, T(17-18: 1), P. *Geometry and Spectra of Compact Riemann Surfaces*. Peter Buser. Progress in Math., V. 106. Birkhäuser, 1992, xiv + 454 pp, \$69.50. [ISBN 0-8176-3406-1] First half introduces geometry of compact Riemann surfaces of genus greater than one based on hyperbolic geometry, cutting and pasting. Includes Fenchel-Nielsen, trigonometry, Bers' partition theorem, and Teichmüller space. Second half introduces the Laplacian's spectrum on such surfaces, and how the spectrum reflects their geometry. SP

Geometry, T*(15-16), L. *The Poincaré Half-Plane: A Gateway to Modern Geometry*. Saul Stahl. Jones & Bartlett, 1993, xiii + 298 pp, \$41.75. [ISBN 0-86720-298-X] An excellent text on modern geometry. Mostly on hyperbolic geometry in the upper half-plane, but also nicely overviews contrast between euclidean and non-euclidean geometry. Many exercises. MPR

Optimization, T(16-17), L. *Introduction to the Calculus of Variations*. Hans Sagan. Dover, 1992, xvi + 449 pp, \$12.95 (P). [ISBN 0-486-67366-9] A re-issue of the 1969 McGraw-Hill text (TR, August-September 1970; Extended Review, February 1971). Mathematically rigorous introduction. First and second variations are thoroughly examined. The many and useful exercises treat both theoretical considerations and computationally feasible examples. SM

Programming, T(13-14: 1). *Programming with Standard ML*. Colin Myers, Chris Clack, Ellen Poon. Prentice Hall, 1993, x + 301 pp. [ISBN 0-13-722075-8] A practical introduction to structured programming using a functional subset of Standard ML. AO

Computer Systems, P. *Proceedings: 7th Annual X Technical Conference*. Ed: Adrian Nye. X Resource, Issue 5. O'Reilly & Assoc, 1993, 272 pp, \$22.50 (P). [ISBN 1-56592-020-1]

Artificial Intelligence, T(15-16: 2), C. *Practical Neural Network Recipes in C++*. Timothy Masters. Academic Pr, 1993, xviii + 493 pp, \$44.95 (P). [ISBN 0-12-479040-2] A "cook-book" introduction to neural networks; assumes no special background. Stresses three-layer feed forward model. Local minimization of the error function is addressed via simulated annealing and genetic optimization. MPR

Computer Science, T(14: 2), C. *Going from C to C++*. Robert J. Traister. Academic Pr, 1993, xiii + 188 pp, \$34.95 (P).

[ISBN 0-12-697412-8] A nuts-and-bolts introduction to C++ moves the reader briskly to the point of writing object-oriented code. Appropriate for anyone already familiar with object-oriented programming, although conceptual differences between procedural and object-oriented languages are given little attention. With disk. MPR

Applications (Engineering), T(15-18: 4). *Advanced Engineering Mathematics*. Dennis G. Zill, Michael R. Cullen. PWS-Kent, 1992, xv + 1200 pp. [ISBN 0-534-92800-5] Six modules: a standard ODE text; linear algebra and vector calculus; systems of differential equations; Fourier series and boundary-value problems; numerical analysis; complex analysis. Answers to odd-numbered problems. SP

Applications (Fluid Dynamics), P. *Annual Review of Fluid Mechanics, Volume 25, 1993*. Eds: John L. Lumley, Milton Van Dyke, Helen L. Reed. Annual Reviews, 1993, x + 641 pp, \$44. [ISBN 0-8243-0725-9]

Applications (Physics), P, L. *Wulff Construction: A Global Shape from Local Interaction*. R. Dobrushin, R. Kotecký, S. Shlosman. Transl. of Math. Mono., V. 104. AMS, 1992, ix + 204 pp, \$130. [ISBN 0-8218-4563-2] A simple question with a complicated answer: What is the shape of a droplet? Authors use Wulff's construction in a long proof of one theorem on the asymptotic shape of a two-dimensional droplet. MPR

Applications (Physics), T(16-18: 1-3), S, P, L. *Principles of Physical Cosmology*. P.J.E. Peebles. Ser. in Physics. Princeton Univ Pr, 1993, xviii + 718 pp, \$29.95 (P). [ISBN 0-691-01933-9] A comprehensive overview and a genuine *tour-de-force*. First section, the Development of Physical Cosmology, covers basic topics: the expanding universe, background radiation, steady state cosmology, etc. Next section, General Relativity and Cosmology, is more technical but still readable. Final section, Topics in Modern Cosmology, covers several more esoteric areas. Chapter introductions make the topics accessible to anyone with a good background in undergraduate physics. MU

Reviewers

LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; CEC: Clifton E. Corzatt, St. Olaf; SM: Steve McKelvey, St. Olaf; AO: Arnold Ostebee, St. Olaf; SP: Samuel Patterson, Carleton; MPR: Matthew P. Richey, St. Olaf; KES: Kay E. Smith, St. Olaf; MU: Milton Ulmer, Carleton; TAV: Theodore A. Vessey, St. Olaf; PZ: Paul Zorn, St. Olaf.

EXCURSIONS IN CALCULUS: an Interplay of the Continuous and the Discrete

Robert M. Young

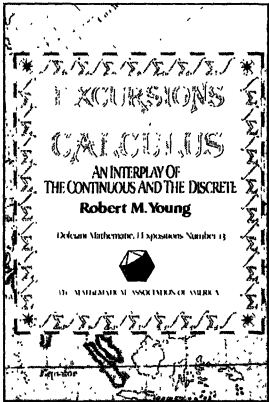
An excellent source of projects for well motivated students. This list of 463 references is a valuable aid for those who wish to dig deeper. —CHOICE

The presentation is clear and the topics very interesting...fully accessible to students for whom the book is intended. The book will be influential in awakening students' awareness for good classical mathematics. —Paulo Ribenboim

Printed with eight full-color plates.

The purpose of this book is to explore, within the context of elementary calculus, the rich and elegant interplay that exists between the two main currents of mathematics, the continuous and the discrete. Such fundamental notions in discrete mathematics as induction, recursion, combinatorics, number theory, discrete probability, and the algorithmic point of view as a unifying principle are continually explored as they interact with traditional calculus. The interaction enriches both.

The book is addressed primarily to well-trained calculus students and their teachers, but it can serve as a supplement in a traditional calculus course for anyone who wants to see more.



CONTENTS:

- Infinite Ascent, Infinite Descent: The Principle of Mathematical Induction
- Patterns, Polynomials, and Primes: Three Applications of the Binomial Theorem
- Fibonacci Numbers: Function and Form
- On the Average
- Approximation: from Pi to the Prime Number Theorem
- Infinite Sums: A Potpourri

The problems, taken for the most part from probability, analysis and number theory, are an integral part of the text. Many point the reader toward further excursions. There are over 400 problems presented in this book.

408 pp., 1992, Paperbound
ISBN 0-88385-317-5
List: \$39.00 MAA Member: \$31.00
Catalog Number DOL-13

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

Membership Code	Qty.	Catalog Number	Price

Name _____			
Address _____			
City _____			
State ____ Zip Code _____			
			Total \$ _____
			Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
			Credit Card No. _____
			Signature _____ Exp. Date _____

TWO OF THE GREATEST

FOUR MORE GREATS

1 **Calculus with Analytic Geometry, 5/e** 0-03-096800-3 1994

Robert Ellis and Denny Gulick both of
the *University of Maryland, College Park*

This bestseller now combines early coverage of logarithmic and exponential functions with trigonometric functions. Its reputation for precise, error-free mathematics is enhanced by the addition of graphing calculator exercises and *Topics for Discussion* which develop students' critical thinking/writing and group skills.

2 **Calculus from Graphical, Numerical, and Symbolic Points of View**

Vol. I (1994 Prelim. Ed.) 0-03-098731-8
Vol. II (1994 Prelim. Ed.) 0-03-098732-6

Arnold Ostebee and Paul Zorn both of
St. Olaf College

Students are challenged to compare graphical, numerical and symbolic viewpoints of calculus. The text uses an early transcendental approach and provides more extensive coverage of differential equations. Appropriate use of technology — graphing calculators or computers — is stressed.



Saunders College Publishing / a division of Harcourt Brace College Publishers
Public Ledger Building / 620 Chestnut Street, Suite 560 / Philadelphia, PA 19106-3477

3 **Elementary Linear Algebra, 5/e** 0-03-097354-6 1994

Stanley I. Grossman, *University of
Montana and University College, London*

This new edition achieves a unique balance between theory and technique, and features Matlab® applications, and an emphasis on geometric interpretation to make linear algebra easier for students to grasp.

4 **Numerical Analysis** 0-03-098330-4 1994

Vithal A. Patel, *Humboldt State
University*

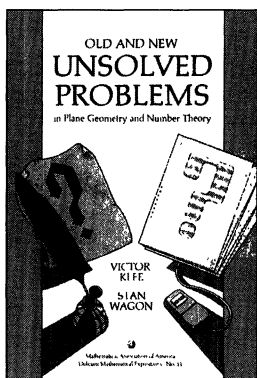
Patel introduces methods and algorithms required to succeed in scientific computing. This new text provides a solid theoretical foundation in numerical analysis enabling students to solve problems successfully and stimulate their interest in developing new numerical methods. Carefully developed computer exercises help students make the transition to writing their own computer problems.

**To order examination copies of
these texts, contact your
Saunders Representative or
call 1-708-647-8822.**

Absolute Values in Math

OLD AND NEW UNSOLVED
PROBLEMS IN PLANE
GEOMETRY AND
NUMBER THEORY

Victor Klee and Stan Wagon



Many facts and problems to fascinate both on familiar and unfamiliar topics. It is compulsive reading and will fill your mind with problems that will come back to haunt you again and again during idle moments.

—Mathematical Intelligencer

The book will serve well as a point of entry for students who want to know more about celebrated questions, or simply take in the vistas.

—CHOICE

This is a book that not only belongs in every university, college and high school library, it very definitely belongs in every public library.

—Mathematical Reviews

Part of the broad appeal of mathematics is that there are simply stated questions that have not yet been answered. These questions are plentiful in the areas of plane geometry and number theory, and the purpose of this book is to discuss some unsolved problems in these fields. Because the central concepts of geometry and number theory are understood by everyone, many of the questions can be understood by readers with extremely little mathematical background.

The authors place each problem in its historical and mathematical context. Each problem section is presented in two parts: The first gives an

elementary overview discussing the history and both solved and unsolved variants of the problem. Part Two contains more details, including a few proofs of related results, a wider and deeper survey of what is known about the problem and its relatives, and a large collection of references. Both parts contain exercises, and solutions to the exercises are included.

The book is aimed at both teachers and students of undergraduate mathematics, and at beginning graduate students. It could be used as a text in a course about unsolved problems, and also in courses in geometry or number theory. High school teachers interested in learning about developments in modern mathematics will find much of interest here.

352 pp., Paperbound, 1991
ISBN 0-88585-315-9
List: \$26.00 MAA Member: \$19.00
Catalog Number DOL-11

ORDER FROM:

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC. 20036
1-800-331-1622 (FAX) (202) 265-2384

Membership Code _____	Quantity	Title	Price
Name _____	_____		
Address _____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
City _____ State _____ Zip _____	Credit Card No. _____	TOTAL \$ _____	
	Signature _____	Exp. Date _____	

NEW FROM MCGRAW-HILL FOR 1994

A MATHEMATICAL JOURNEY, Second Edition

Stanley Gudder, *University of Denver*

ELEMENTARY ALGEBRA:

STRUCTURE AND Use, Sixth Edition

INTERMEDIATE

ALGEBRA: STRUCTURE AND Use, Fifth Edition

both by Raymond Barnett,
Merritt College

Thomas Kearns, *Northern
Kentucky University*

PREALGEBRA, First Edition

ESSENTIAL MATHEMATICS WITH APPLICATIONS, Second Edition

both by Lawrence A. Trivieri, *DeKalb College*

FINITE MATHEMATICS AND ITS APPLICATIONS, Second Edition

Stanley J. Farlow, *University of Maine*

DISCOVERING CALCULUS: A PRELIMINARY VERSION, Volumes I & II

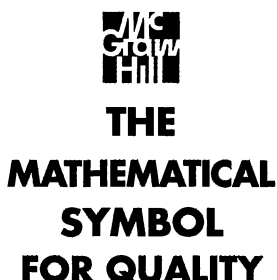
Alan L. Levine & George M. Rosenstein, Jr.,
Both of Franklin and Marshall College

FINITE MATHEMATICS WITH COMPUTER-SUPPORTED APPLICATIONS

James Calvert & William Voxman,
Both of The University of Idaho

CONTEMPORARY STATISTICS: A COMPUTER APPROACH

Sheldon P. Gordon,
Suffolk Community College
Florence S. Gordon,
New York Institute of Technology



AN INTRODUCTION TO DIFFERENTIAL EQUATIONS AND THEIR APPLICATIONS

Stanley J. Farlow, *University of Maine*

ENGINEERING MATHEMATICS WITH MATHEMATICA

John S. Robertson, *U.S. Military Academy*

Also Available

CALCULUS AND ANALYTIC GEOMETRY, 5/e

Sherman K. Stein,
University of California, Davis
Anthony Barcellos,
American River College

CALCULUS LABORATORIES WITH MATHEMATICA, VOLS. I, II, III

Michael Kerckhove and Van Nall,
Both of University of Richmond

DISCOVERING CALCULUS WITH THE TI-S1 AND THE TI-85

DISCOVERING CALCULUS WITH THE CASIO fx-7700 AND fx-8700

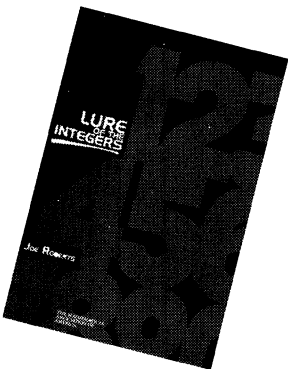
DISCOVERING CALCULUS WITH THE HP-28 AND THE HP-48

Robert T. Smith,
Millersville University of Pennsylvania
Roland B. Minton,
Roanoke College

For more information, please contact your local McGraw-Hill representative or write: McGraw-Hill College Division, Comp Processing & Control, P.O. Box 448, Hightstown, New Jersey 08520-0448.

LURE OF THE INTEGERS

Joe Roberts



A joy to read and ponder, this book is a welcome addition to the body of mathematical literature. It belongs in every mathematical library.
—*Journal of Recreational Mathematics*

Will enrich library collections serving curricula with theory of numbers courses. —*Choice*

In some small way, this book is an introduction to a mythical book which might go under the name of *The Book of Integers*. This mythical book has on page n all of the interesting properties of the integer n . This introduction stems from many years' casual accumulation of numerical facts. Most of the material presented belongs to elementary mathematics in the sense that no deep or profound mathematical background is required in order to understand what is said. Much of the material is drawn from the theory of numbers.

Many of the topics touch on contemporary research and most of the results are stated without proof. As a general rule, one cannot tell from the statements of the results whether or not their proofs will be elementary. Indeed, this is a hallmark of mathematics and is one of the things that gives the subject a special flavor and interest. Until one knows that expert practitioners have been unable to solve a problem, one does not know that the problem is difficult. Even then it may turn out that there is an easy solution.

Some of the material will be familiar to people having only a small acquaintance with mathematics. Even in those cases, the author provides something new. On the other hand, much of the material is sufficiently out of the main stream of concern that even professional mathematicians may be unfamiliar with the results. The many references to the literature will almost always enable a reader to track down further information. In **Lure of the Integers** the author has presented a body of material which will prove interesting to the enlightened layman as well as to the professional.

300 pp., Paperbound, 1992
ISBN-0-88385-502-X
List: \$28.50 MAA Member: \$19.50
Catalog Number LURE

ORDER FROM:
The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

Foreign Orders Please add \$3.00 per item ordered to cover postage and handling fees. The order will be sent via surface mail. If you want your order sent by air, we will be happy to send you a proforma invoice for your order.

Membership Code _____
Name _____
Address _____
City _____
State _____ Zip Code _____

Qty.	Catalog Number	Price
_____	_____	_____
_____	_____	_____
		Total \$ _____
Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
Credit Card No. _____		
Signature _____		Exp. Date _____

Symbolic Computation in Undergraduate Mathematics Education

Zaven Karian, Editor

If you are considering putting a symbolic computing system into your curriculum, this is one publication you should have.

—Mathematics Teacher

This well-written book should be helpful to anyone using symbolic computation as an aid in teaching undergraduates—The book provides a number of examples for presenting probability and statistics in a way that removes the tedium and emphasizes the underlying ideas.

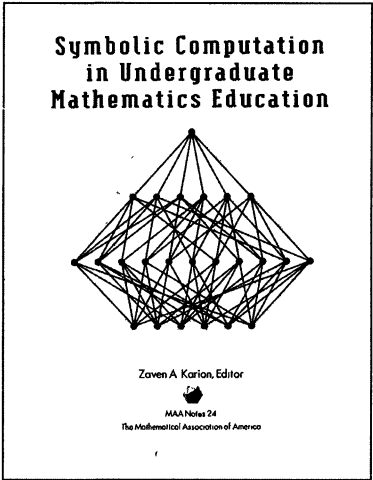
—AAAS, Science Book and Films

If you have any plans to integrate symbolic computing into your program, read and study this book first. Your students will thank you for it.

—AMATYC Review

This volume brings together many of the facets associated with the pedagogic uses of symbolic computation.

Part I consists of articles that deal with general issues of learning mathematics and the role of symbolic computation in that process. The articles in Part II describe the use of symbolic computation in teaching calculus. Some of the areas covered are the use of symbolic computation in a laboratory calculus course, the uses of Derive in the instruction of calculus, antidifferentiation and the



definite integral, and the experiences and reflections of teachers who have used symbolic computation in calculus instruction.

Part III consists of papers on sophomore-level courses on linear algebra and differential equations. The articles in Part IV describe what can be done in using symbolic computation in teaching combinatorics, probability and statistics courses. The articles and references in Part V will help you get started in using some of these ideas at your own institution.

200 pp., 1992, Paperbound

ISBN 0-88385-082-6

List: \$22.00

Catalog Number NTE-24

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 2003
1-800-331-1622 Fax (202) 265-2384

Membership Code

Name _____
Address _____
City _____
State _____ Zip Code _____

Qty.	Catalog Number	Price
_____	_____	_____
_____	_____	_____
		Total \$ _____
Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
Credit Card No. _____		
Signature _____ Exp. Date _____		

Mathematical Cranks

Underwood Dudley

A delightful collection...It is hard to put down and provides topics for an unending series of interesting discussions. The organization and breadth of the book are impressive, supported by a helpful index and a list of resources that encourage further explorations. A classic. —CHOICE

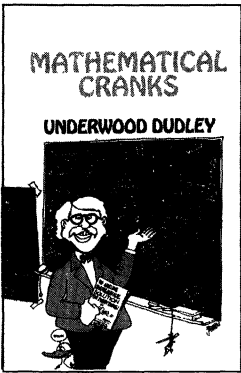
A jewel...The most interesting book that I have read this year.

—Journal of Recreational Mathematics

It's a gem...Dudley hasn't tossed out crank submissions over the years; he's saved them and collected samples from other mathematicians. And what wonderful samples they are.

—Sunday Telegraph, Nashua, New Hampshire

Mathematical Cranks is about people who think that they have done something impossible, like trisecting the angle, squaring the circle, duplicating the cube, or proving Euclid's parallel postulate; people who think they have done something that they have not, like proving Fermat's Last Theorem, verifying Goldbach's Conjecture, or finding a simple proof of the Four Color Theorem; people who have eccentric views, from mild (thinking we should count by 12s instead of 10s) to crazy (thinking that second-order differential equations will solve all problems of economics, politics, and philosophy); people who pray in matrices; people who find the American Revolution ruled by the number 57; people who have in common something to do with mathematics and something odd, peculiar, or bizarre.



Cranks and their ideas come in great variety. The book is a collection of examples, designed to give readers an idea of what cranks do and how they do it. Contemplating the odd, peculiar, or bizarre can be entertaining or enlightening. There can be no solution to the problem of mathematical cranks—obsessive people we will always have with us, and some will become obsessed with mathematics—but perhaps viewing the futility of their efforts will turn some prospective cranks toward more fruitful endeavors.

This is a truly unique book, written with wit and style.

300 pp., 1992, Paperbound
ISBN 0-88385-507-0
List: \$25.00 MAA Member: \$18.00
Catalog Number CRANKS


ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

Membership Code	Qty.	Catalog Number	Price

Name _____			
Address _____			
City _____			
State _____ Zip Code _____			
			Total \$ _____
			Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
			Credit Card No. _____
			Signature _____ Exp. Date _____

TOOLBOOKS



The AP PROFESSIONAL
ToolKit—We give you
the tools, the solutions,
the references...
Right at your fingertips.

The Alternative to Binary Logic...

The Fuzzy Systems Handbook

A Practitioner's Guide to
Building, Using, and
Maintaining Fuzzy Systems

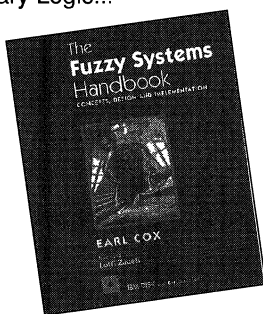
Earl Cox

Foreword by
Loffi Zadeh

February 1994

Paperback, \$49.95, c. 630 pp.
ISBN: 0-12-194270-8

Includes one IBM disk with C++ source code.



INTRODUCTION TO COMPUTER PERFORMANCE ANALYSIS WITH MATHEMATICA®

Arnold O. Allen
Hewlett-Packard Company

October 1993

Hardcover, \$49.95, 384 pp.
ISBN: 0-12-051070-7

Includes one 3 1/2" disk.



Mathematica® By Example

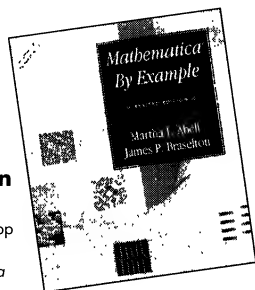
REVISED EDITION

Martha L. Abell
James P. Braselton

January 1994

Paperback, \$39.95, c. 544 pp.
ISBN: 0-12-041530-5

Compatible with Mathematica
Version 2.2.



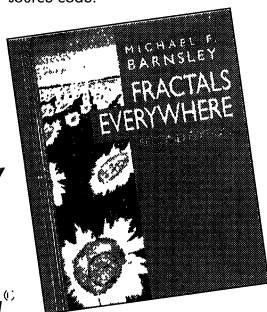
FRACTALS EVERYWHERE

SECOND EDITION

Michael F. Barnsley

August 1993

Hardcover, \$49.95, 531 pp.
ISBN: 0-12-079061-0



The Mathematica® Programmer

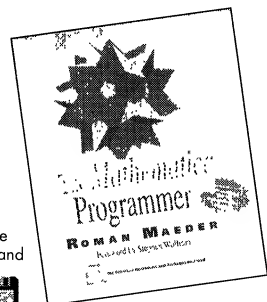
Roman E. Maeder

Foreword by
Stephen Wolfram

December 1993

Paperback, \$44.95, 216 pp.
ISBN: 0-12-464990-4

Includes one disk containing
Mathematica Notebooks and
Packages, which are compatible
with Mathematica Version 2.2 and
its programming language.



Differential Equations with Mathematica®

Martha L. Abell
James P. Braselton

October 1993

Paperback, \$44.95, 631 pp.
ISBN: 0-12-041539-9

The Mathematica® Handbook

Martha L. Abell
James P. Braselton

October 1993

Paperback, \$39.95, 789 pp.
ISBN: 0-12-041536-4

Mastering Mathematica®

Programming Methods and Applications

John W. Gray

January 1994

Paperback, \$44.95, c. 400 pp.
ISBN: 0-12-296040-8

Includes one disk containing Mathematica Notebooks.

Mathematica is a registered trademark of
Wolfram Research, Inc.



AVAILABLE FROM YOUR
LOCAL BOOKSHELF



1-800-321-5068

e-mail: app@acad.com

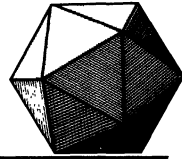
Prices are subject to change without notice.

©AP PROFESSIONAL 1994 RYAN/LMW/ST 07024 12/93



The American Mathematical Monthly

Volume 101 Number 2 / FEBRUARY 1994
(ISSN 0002-9890)



Contents

ARTICLES

Yueh-Gin Gung and Dr. Charles Y. Hu Award for Distinguished Service to
J. Sutherland Frame / DAVID W. BALLEW 107

A New Look at Euler's Theorem for Polyhedra / BRANKO GRÜNBAUM
and G. C. SHEPHARD 109

Otto Neugebauer: Reminiscences and Appreciation / PHILIP J. DAVIS
129

From the Buffon Needle Problem to the Kreiss Matrix Theorem /
ELIAS WEGERT and LLOYD N. TREFETHEN 132

A Counterexample for Germain / WILLIAM C. WATERHOUSE 140

Cubic Equations, or Where Did the Examination Question Come From?/
H. B. GRIFFITHS and A. E. HIRST 151

FEATURES

COMMENTS 106

PICTURE PUZZLE 139

NOTES

On the Identity of Polyhedra / HELLMUTH STACHEL 162

More on the Pompeiu Problem / DAVID C. ULLRICH 165

UNSOLVED PROBLEMS

Every Number Is Expressible as the Sum of How Many
Polygonal Numbers? / RICHARD K. GUY 169

THE AUTHORS 173

PROBLEMS AND SOLUTIONS 175

REVIEWS

Reality Rules I. The Fundamentals; II. The Frontier. By John Casti /
RUTHERFORD ARIS 186

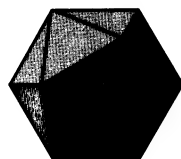
Geometry of Surfaces. By John Stillwell / DAVID L. WEBB 188

TELEGRAPHIC REVIEWS 191

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.



The American Mathematical Monthly



Volume 101, Number 3 / MARCH 1994



NOTICE TO AUTHORS

The *Monthly* publishes articles, notes, and other features about mathematics and the profession. The readership of the *Monthly* is intended to include everybody who is mathematically inclined, including of course professional mathematicians and students of mathematics at all collegiate levels. While no single article or feature is likely to appeal to everyone, material should interest and be accessible to a large number of readers. This is the most important criterion for acceptance.

Articles may be expositions of old results or presentations of new ones. They may concern all of mathematics or one small area, a broad development or a single application, historical reminiscences or one important event. While some articles may contain the author's new research, the novelty of material and generality of the results is far less important than the clarity of exposition and general interest. Discussing one illuminating case of a well known result is far better than providing all the details of an obscure but new proposition. Articles in the *Monthly* are supposed to inform and to entertain; they are meant to be read rather than archived.

Notes are short and possibly informal articles. A note may concern a clever new proof of an old theorem, a novel way to present tired material, or a lively discussion of a philosophical (but still mathematical) issue. Also, any topic is suitable, so long as it is related to mathematics. Because a note is short, the first few sentences are the most important part: They should explain the purpose and invite the reader in. Photographs or diagrams often will attract the reader's attention.

All articles and notes should be sent to the editor:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Please send 3 copies, typewritten on only one side of the paper. Illustrations should be carefully drawn on separate sheets of paper in black ink; the original should be without lettering and two copies should have appropriate captions and lettering indicated.

Proposed problems or solutions should be sent to:

RICHARD BUMBY,
P.O. Box 10971
New Brunswick, NJ 08906-0971.

Please send 2 copies of all material, typewritten if possible.

Letters to the Editor, both for publication and for private reading, should be sent to the Editor at the address given above. Comments, including criticisms, are welcome, as are all suggestions for making the *Monthly* a lively, entertaining, and informative journal.

EDITOR:

JOHN H. EWING

ASSOCIATE EDITORS:

RONALD BOOK	FRED KOCHMAN
PETER BORWEIN	CATHERINE MCGEOCH
RICHARD BUMBY	RICHARD NOWAKOWSKI
DENNIS DETURCK	ARNOLD OSTELEE
UNDERWOOD DUDLEY	LEE RUBEL
JOHN DUNCAN	ABE SHENITZER
JOAN FERRINI-MUNDY	LYNN STEEN
JOSEPH GALLIAN	STAN WAGON
STEVEN GALOVICH	DOUGLAS WEST
RICHARD GUY	HERBERT WILF
DARRELL HAILE	SANDY ZABELL
PAUL HALMOS	PAUL ZORN
JOAN HUTCHINSON	

EDITORIAL ASSISTANT:

MISTY CUMMINGS

STAFF ARTIST:

MIKE CAGLE

Reprint permission:

MARCIA P. SWARD, Executive Director

Advertising Correspondence:

Ms. ELAINE PEDREIRA, Advertising Manager

Subscription correspondence, change of address, and other inquiries:

Membership / Subscriptions Department

All at the address:

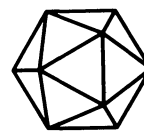
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT. Copyrighted by the Mathematical Association of America (Incorporated), 1994, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source. Second class postage paid at Washington, DC, and additional mailing offices. **Postmaster:** Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.

**The American
Mathematical Monthly**

Volume 101 Number 3 / MARCH 1994
(ISSN 0002-9890)



Contents

ARTICLES

- A Marvelous Proof / FERNANDO Q. GOUVÊA 203
Triangulating the Circle, at Random / DAVID ALDOUS 223
Hypatia and Her Mathematics / MICHAEL A. B. DEAKIN 234
Calculus II and Euler Also (with a Nod to Series Integral
Remainder Bounds) / RICHARD BARSHINGER 244
A Focusing Property of the Ellipse / MARC FRANTZ 250
-

FEATURES

COMMENTS 202

NOTES

- Reflections Can Be Trapped / ROBERTO PEIRONE 259
Euler's Theorem / KATHERINE HEINRICH
and PETER HORAK 260

PICTURE PUZZLE 261

THE COMPUTER SCIENCE SAMPLER

- Universal Traversal Sequences / JOAN FEIGENBAUM
and NICK REINGOLD 262

THE EVOLUTION OF...

- What Are Algebraic Integers and What Are They For? /
JOHN STILLWELL 266

THE AUTHORS 271

PROBLEMS AND SOLUTIONS 273

REVIEWS

- Revolutions in Mathematics*. Edited by Donald Gillies /
MICHAEL MAHONEY 283

TELEGRAPHIC REVIEWS 288

COMMENTS

"In view of the speculation on the status of my work on the Taniyama-Shimura conjecture and Fermat's Last Theorem I will give a brief account of the situation. During the review process a number of problems emerged, most of which have been resolved, but one in particular I have not yet settled. The key reduction of (most cases of) the Taniyama-Shimura conjecture to the calculation of the Selmer group is correct. However the final calculation of a precise upper bound for the Selmer group in the semistable case (of the symmetric square representation associated to a modular form) is not yet complete as it stands. I believe that I will be able to finish this in the near future using the ideas explained in my Cambridge lectures.

"The fact that a lot of work remains to be done on the manuscript makes it still unsuitable for release as a preprint. In my course in Princeton beginning in February I will give a full account of this work."

Andrew Wiles (December, 1993).

Well, this is a fine mess. When word flashed around the world last summer that Andrew Wiles had proved the Fermat Conjecture, mathematicians quickly took advantage of the spotlight. Articles about *mathematics* not only appeared in Science and Science News, they also appeared in Time magazine and the Bloomington Herald Times. Eric Zorn of the Chicago Tribune wrote a humorous column about the solution. Marilyn vos Savant wrote an article in Parade Magazine. (The proof used "hyperbolic geometry", she said, so it was questionable.) Important people (mainly in Washington) pointed out that this was a triumph of "useless" mathematics, which was supported by public funds for over 40 years (and we need more). The rest of us, concerned with funding on a smaller scale, sold T-shirts with Fermat on the front.

The excitement had a downside for editors. The usual trickle of Fermat proofs became a torrent. Covering letters scoffed at a proof in 200 pages; here's one in four! And the excitement encouraged those who trisect angles and square circles and prove well known results on the back of a postcard. Rejection letters for such manuscripts are polite but firm.

Now what do I tell those people? If Andrew Wiles made a mistake, why not the editor as well? What do we tell the readers of Time who thought mathematics dealt in absolute truths? What do we tell the congressman, who was ready to double funding for mathematics? (Well, maybe just for Number Theory.)

We were carried away by the excitement. Andrew Wiles warned us --- the proof wasn't written down yet. Heady with public attention, we didn't listen.

The January issue of the Monthly contained an article by David Cox on the Fermat Conjecture and its history. In this issue we publish another article on the mathematics *behind* the recent work of Wiles. Why another? Fernando Gouvêa wrote this article soon after the announcement by Wiles last summer, well before anyone could check the details. Its purpose is to explain some of the lovely mathematics that underlies Wiles's attempted proof. In the coming months, as mathematicians debate the proof and its status, this mathematics becomes even more important. Learning some of the terms and ideas will make us better able to follow that debate . . . and perhaps more cautious about sensational pronouncements in the future.

John Ewing

“A Marvelous Proof”

Fernando Q. Gouvêa

No one really knows when it was that the story of what came to be known as “Fermat’s Last Theorem” really started. Presumably it was sometime in the late 1630s that Pierre de Fermat made that famous inscription in the margin of Diophantus’ *Arithmetic* claiming to have found “a marvelous proof”. It seems now, however, that the story may be coming close to an end. In June, 1993, Andrew Wiles announced that he could prove Fermat’s assertion. Since then, difficulties seem to have arisen, but Wiles’ strategy is fundamentally sound and may yet succeed.

The argument sketched by Wiles is an artful blend of various topics that have been, for years now, the focus of intensive research in number theory: elliptic curves, modular forms, and Galois representations. The goal of this article is to give mathematicians who are not specialists in the subject access to a general outline of the strategy proposed by Wiles. Of necessity, we concentrate largely on background material giving first a brief description of the relevant topics, and only afterwards describe how they come together and relate to Fermat’s assertion. Readers who are mainly interested in the structure of the argument and who do not need or want too many details about the background concepts may want to skim through Section 2, then concentrate on Section 3. Our discussion includes a few historical remarks, but history is not our main intention, and therefore we only touch on a few highlights that are relevant to our goal of describing the main ideas in Wiles’ attack on the problem.

Thanks are due to Barry Mazur, Kenneth Ribet, Serge Lang, Noriko Yui, George Elliot, Keith Devlin, and Lynette Millett for their help and comments.

1 PRELIMINARIES. We all know the basic statement that Fermat wrote in his margin. The claim is that for any exponent $n \geq 3$ there are no non-trivial integer solutions of the equation $x^n + y^n = z^n$. (Here, “non-trivial” will just mean that none of the integers x , y , and z is to be equal to zero.) Fermat claims, in his marginal note, to have found “a marvelous proof” of this fact, which unfortunately would not fit in the margin.

This statement became known as “Fermat’s Last Theorem,” not, apparently, due to any belief that the “theorem” was the last one found by Fermat, but rather due to the fact that by the 1800s all of the other assertions made by Fermat had been either proved or refuted. This one was the last one left open, whence the name. In what follows, we will adopt the abbreviation FLT for Fermat’s statement, and we will refer to $x^n + y^n = z^n$ as the “Fermat equation.”

The first important results relating to FLT were theorems that showed that Fermat’s claim was true for specific values of n . The first of these is due to Fermat himself: very few of his proofs were ever made public, but in one that was he shows

that the equation

$$x^4 + y^4 = z^2$$

has no non-trivial integer solutions. Since any solution of the Fermat equation with exponent 4 gives a solution of the equation also, it follows that Fermat's claim is true for $n = 4$.

Once that is done, it is easy to see that we can restrict our attention to the case in which n is a prime number. To see this, notice that any number greater than 2 is either divisible by 4 or by an odd prime, and then notice that we can rewrite an equation

$$x^{mk} + y^{mk} = z^{mk}$$

as

$$(x^m)^k + (y^m)^k = (z^m)^k,$$

so that any solution for $n = mk$ yields at once a solution for $n = k$. If n is not prime, we can always choose k to be either 4 or an odd prime, so that the problem reduces to these two cases.

In the 1750s, Euler became interested in Fermat's work on number theory, and began a systematic investigation of the subject. In particular, he considered the Fermat equation for $n = 3$ and $n = 4$, and once again proved that there were no solutions. (Euler's proof for $n = 3$ depends on studying the "numbers" one gets by adjoining $\sqrt{-3}$ to the rationals, one of the first instances where one meets "algebraic numbers.") A good historical account of Euler's work is to be found in [Wei83]. In the following years, several other mathematicians extended this step by step to $n = 5, 7, \dots$. A general account of the fortunes of FLT during this time can be found in [Rib79].

Since then, ways for testing Fermat's assertion for any specific value of n have been developed, and the range of exponents for which the result was known to be true kept getting pushed up. As of 1992, one knew that FLT was true for exponents up to 4 000 000 (by work of J. Buhler).

It is clear, however, that to get general results one needs a general method, i.e., a way to connect the Fermat equation (for any n) with some mathematical context which would allow for its analysis. Over the centuries, there have been many attempts at doing this; we mention only the two biggest successes (omitting quite a lot of very good work, for which see, for example, [Rib79]).

The first of these is the work of E. Kummer, who, in the mid-nineteenth century, established a link between FLT and the theory of cyclotomic fields. This link allowed Kummer to prove Fermat's assertion when the exponent was a prime that had a particularly nice property (Kummer named such primes "regular"). The proof is an impressive bit of work, and was the first general result about the Fermat equation. Unfortunately, while in numerical tests a good percentage of primes seem to turn out to be regular, no one has yet managed to prove even that there are infinitely many regular primes. (And, ironically, we do have a proof that there are infinitely many primes that are *not* regular.) A discussion of Kummer's approach can be found in [Rib79]; for more detailed information on the cyclotomic theory, one could start with [Was82].

The second accomplishment we should mention is that of G. Faltings, who, in the early 1980s, proved Mordell's conjecture about rational solutions to certain kinds of polynomial equations. Applying this to the Fermat equations, one sees that for any $n \geq 4$ one can have only a *finite* number of non-trivial solutions. Once

again, this is an impressive result, but its impact on FLT itself turns out to be minor because we have not yet found a way to actually determine how many solutions should exist. For an introduction to Faltings' work, check [CS86], which contains an English translation of the original paper.

Wiles' attack on the problem turns on another such linkage, also developed in the early 1980s by G. Frey, J.-P. Serre, and K. A. Ribet. This one connects FLT with the theory of elliptic curves, which has been much studied during all of this century, and thereby to all the machinery of modular forms and Galois representations that is the central theme of Wiles' work. The main goal of this paper is to describe this connection and then to explain how Wiles attempts to use it to prove FLT.

Notation. We will use the usual symbols \mathbb{Q} for the rational numbers and \mathbb{Z} for the integers. The integers modulo m will be written¹ as $\mathbb{Z}/m\mathbb{Z}$; we will most often need them when m is a power of a prime number p . If p is prime, then $\mathbb{Z}/p\mathbb{Z}$ is a field, and we commemorate that fact by using an alternative notation: $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$.

2 THE ACTORS. We begin by introducing the main actors in the drama. First, we briefly (and very informally) introduce the p -adic numbers. These are not so much actors in the play as they are part of the stage set: tools to allow the actors to do their job. Then we give brief and impressionistic outlines of the theories of Elliptic Curves, Modular Forms, and Galois Representations.

2.1 p -adic Numbers. The p -adic numbers are an extension of the field of rational numbers which are, in many ways, analogous to the real numbers. Like the real numbers, they can be obtained by defining a notion of distance between rational numbers, and then passing to the completion with respect to that distance. For our purposes, we do not really need to know much about them. The crucial facts are:

1. For each prime number p there exists a field \mathbb{Q}_p which is complete with respect to a certain notion of distance and contains the rational numbers as a dense subfield.
2. Proximity in the p -adic metric is closely related to congruence properties modulo powers of p . For example, two integers whose difference is divisible by p^n are "close" in the p -adic world (the bigger the n , the closer they are).
3. As a consequence, one can think of the p -adics as encoding congruence information: whenever one knows something modulo p^n for every n , one can translate this into p -adic information, and vice-versa.
4. The field \mathbb{Q}_p contains a subring \mathbb{Z}_p , which is called the *ring of p -adic integers*. (In fact, \mathbb{Z}_p is the closure of \mathbb{Z} in \mathbb{Q}_p .)

There is, of course, a lot more to say, and the reader will find it said in many references, such as [Kob84], [Cas86], [Ami75], and even [Gou93]. The p -adic numbers were introduced by K. Hensel (a student of Kummer), and many of the basic ideas seem to appear, in veiled form, in Kummer's work; since then, they have become a fundamental tool in number theory.

¹Many elementary texts like to use \mathbb{Z}_m as the notation for the integers modulo m ; for us (and for serious number theory in general), this notation is inconvenient because it collides with the notation for the p -adic integers described below.

2.2 Elliptic Curves. Elliptic curves are a special kind² of algebraic curves which have a very rich arithmetical structure. There are several fancy ways of defining them, but for our purposes we can just define them as the set of points satisfying a polynomial equation of a certain form.

To be specific, consider an equation of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

where the a_i are integers (there is a reason for the strange choice of indices on the a_i , but we won't go into it here). We want to consider the set of points (x, y) which satisfy this equation. Since we are doing number theory, we don't want to tie ourselves down too seriously as to what sort of numbers x and y are: it makes sense to take them in the real numbers, in the complex numbers, in the rational numbers, and even, for any prime number p , in \mathbb{F}_p (in which case we think of the equation as a congruence modulo p). We will describe the situation by saying that there is an underlying object which we call *the curve* E and, for each one of the possible fields of definition for points (x, y) , we call the set of possible solutions the “points of E ” over that field. So, if we consider all possible complex solutions, we get the set $E(\mathbb{C})$ of the complex points of E . Similarly, we can consider the real points $E(\mathbb{R})$, the rational points $E(\mathbb{Q})$, and even the \mathbb{F}_p -points $E(\mathbb{F}_p)$.

We haven't yet said when it is that such equations define elliptic curves. The condition is simply that the curve be *smooth*. If we consider the real or complex points, this means exactly what one would expect: the set of points contains no “singular” points, that is, at every point there is a well-defined tangent line. We know, from elementary analysis, that an equation $f(x, y) = 0$ defines a smooth curve exactly when there are no points on the curve at which both partial derivatives of f vanish. In other words, the curve will be smooth if there are no common solutions of the equations

$$f(x, y) = 0 \quad \frac{\partial f}{\partial x}(x, y) = 0 \quad \frac{\partial f}{\partial y}(x, y) = 0.$$

Notice, though, that this condition is really algebraic (the derivatives are derivatives of polynomials, and hence can be taken formally). In fact, we can boil it down to a (complicated) polynomial condition in the a_i . There is a polynomial $\Delta(E) = \Delta(a_1, a_2, a_3, a_4, a_6)$ in the a_i such that E is smooth if and only if $\Delta(E) \neq 0$. This gives us the means to give a completely formal definition (which makes sense even over \mathbb{F}_p). The number $\Delta(E)$ is called the *discriminant* of the curve E .

Definition 1. *Let K be a field. An elliptic curve over K is an algebraic curve determined by an equation of the form*

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

where each of the a_i belongs to K and such that $\Delta(a_1, a_2, a_3, a_4, a_6) \neq 0$.

Specialists would want to rephrase that definition to allow other equations, provided that a well-chosen change of variables could transform them into equations of this form.

²Perhaps it's best to dispel the obvious confusion right up front: ellipses are not elliptic curves. In fact, the connection between elliptic curves and ellipses is a rather subtle one. What happens is that elliptic curves (over the complex numbers) are the “natural habitat” of the elliptic integrals which arise, among other places, when one attempts to compute the arc length of an ellipse. For us, this connection will be of very little importance.

It's about time to give some examples. To make things easier, let us focus on the special case in which the equation is of the form $y^2 = g(x)$, with $g(x)$ a cubic polynomial (in other words, we're assuming $a_1 = a_3 = 0$). In this case, it's very easy to determine when there can be singular points, and even what sort of singular points they will be. If we put $f(x, y) = y^2 - g(x)$, then we have

$$\frac{\partial f}{\partial x}(x, y) = -g'(x) \quad \text{and} \quad \frac{\partial f}{\partial y}(x, y) = 2y,$$

and the condition for a point to be “bad” becomes

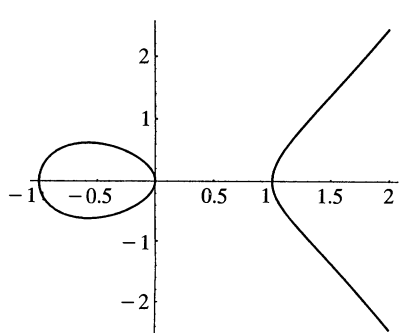
$$y^2 = g(x) \quad -g'(x) = 0 \quad 2y = 0,$$

which boils down to $y = g(x) = g'(x) = 0$. In other words, a point will be bad exactly when its y -coordinate is zero and its x -coordinate is a *double root* of the polynomial $g(x)$. Since $g(x)$ is of degree 3, this gives us only three possibilities:

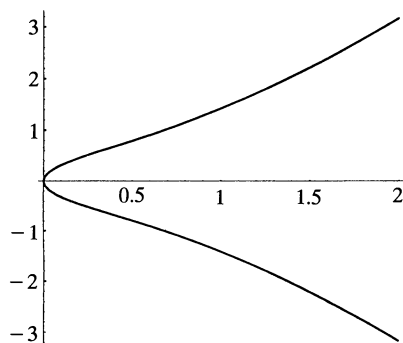
- $g(x)$ has no multiple roots, and the equation defines an elliptic curve;
- $g(x)$ has a double root;
- $g(x)$ has a triple root.

Let's look at one example of each case, and graph the real points of the corresponding curve.

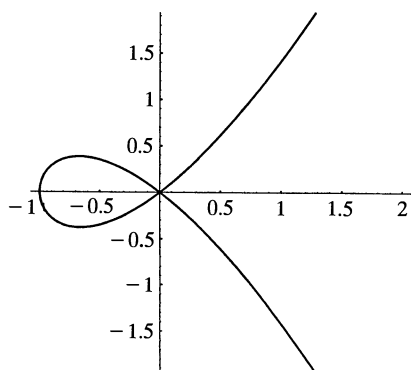
For the first case, consider the curve given by $y^2 = x^3 - x$. Its graph is in figure 1(a) (to be precise, this is the graph of its real points). A different example of the same case is given by $y^2 = x^3 + x$; see figure 1(b). (The reason these look so



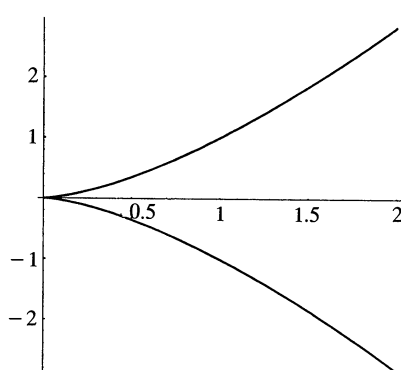
(a) $y^2 = x^3 - x$



(b) $y^2 = x^3 + x$



(c) $y^2 = x^3 + x^2$: a node



(d) $y^2 = x^3$: a cusp

different is that we are only looking at the real points of the curve; in fact, over the complex numbers these two curves are isomorphic.)

When there are “bad” points, what has happened is that either two roots of $g(x)$ have “come together” or all three roots have done so. In the first case, we get a loop. At the crossing point, which is usually called a “node,” the curve has two different tangent lines. See Figure 1(c), where we have the graph of the equation $y^2 = x^3 + x^2$ (double root at zero).

In the final case, not only have all three roots of $g(x)$ come together, but also the two tangents in the node have come together to form a sort of “double tangent” (this can be made precise with some easy algebra of polynomials, but it’s more fun to think of it geometrically). The graph now looks like Figure 1(d), and we call this kind of singular point a “cusp”.

How does all this relate to the discriminant Δ we mentioned above? Well, if r_1, r_2 and r_3 are the roots of the polynomial $g(x)$, the discriminant for the equation $y^2 = g(x)$ turns out to be

$$\Delta = K(r_1 - r_2)^2(r_1 - r_3)^2(r_2 - r_3)^2,$$

where K is a constant. This does just what we want: if two of the roots are equal, it is zero, and if not, not. Furthermore, it is not too hard to see that Δ is actually a polynomial in the coefficients of $g(x)$, which is what we claimed. In other words, all that the discriminant is doing for us is giving a direct algebraic procedure for determining whether there are singular points.

While this analysis applies specifically to curves of the form $y^2 = g(x)$, it actually extends to all equations of the sort we are considering: there is at most one singular point, and it is either a node or a cusp.

One final geometric point: as one can see from the graphs, these curves are not closed. It is often convenient to “close them up.” This is done by adding one more point to the curve, usually referred to as “the point at infinity.” This can be done in a precise way by embedding the curve into the projective plane, and then taking the closure. For us, however, the only important thing is to remember that we actually have one extra point on our curves. (One should imagine it to be “infinitely far up the y -axis,” but keep in mind that there is only one “point at infinity” on the y -axis, so that it is *also* “infinitely far down.”)

With some examples in hand, we can proceed to deeper waters. In order to understand the connection we are going to establish between elliptic curves and FLT, we need to review quite a large portion of what is known about the rich arithmetic structure of these curves.

The first thing to note is that one can define an operation on the set of points of an elliptic curve that makes it, in a natural way, an abelian group. The operation is usually referred to as “addition.” The identity element of this group turns out to be the point at infinity (it would be more honest to say that we *choose* the point at infinity for this role).

We won’t enter into the details of how one adds points on an elliptic curve. In fact, there are several equivalent definitions, each of which has its advantages! The reader should see the references for more details of how it is done (and the proof that one does get a group). The main thing to know about the definition, for now, is that it preserves the field of definition of the points: adding two rational points gives a rational point, and so on.

What this means is that for every choice of a base field, we can get a group of points on the curve with coordinates in that field, so that in fact an elliptic curve gives us a whole bunch of groups, which are, of course, all related (though

sometimes related in a mysterious way). So, given an E , we can look at the complex points $E(\mathbb{C})$, which form a complex Lie group which is topologically a torus, or we can look at the real Lie group $E(\mathbb{R})$, which turns out to be either isomorphic to the circle S^1 or to the direct product $\mathbb{Z}/2\mathbb{Z} \times S^1$. (Look back at the examples above; can you see which is which?)

From an arithmetical point of view, however, the most interesting of these groups is the group of rational points, $E(\mathbb{Q})$. A point $P \in E(\mathbb{Q})$ gives a solution in rational numbers of our cubic equation, and looking for such solutions is, of course, an example of solving a diophantine equation, a sort of problem that is quite important in number theory. What is especially nice about $E(\mathbb{Q})$ is the fact, proved by L. Mordell (and extended by A. Weil) in the 1920s, that it is a *finitely generated* abelian group. What this means is just the following: there is a finite list of rational points on the curve (or, if one prefers, of rational solutions to the equation) such that every other rational solution is obtained by combining (using the addition law) these points with one another. These points are called the *generators* of the group $E(\mathbb{Q})$, which is usually called the *Mordell-Weil group* of E .

The curves we considered earlier have very simple Mordell-Weil groups. For the curve given by $y^2 = x^3 - x$ (figure 1a), it has four points; and for $y^2 = x^3 + x$ (figure 1b) it has two. It is easy, though, to give more interesting examples. Here is one, chosen at random from [Cre92]: if E is the curve defined by $y^2 + y = x^3 - x^2 - 2x + 2$, the Mordell-Weil group $E(\mathbb{Q})$ is an infinite cyclic group, generated by the point $(2, 1)$.

Of course, knowing that we have a finitely generated group raises the obvious question of estimating or computing the number of generators needed and of how one might go about actually finding these generating points. Both of these questions are still open, even though there are rather precise conjectures about what their answers should be. For many specific curves, both the number and the generators themselves have been completely worked out (see, for example, the tables in [Cre92]), but the general problem still seems quite difficult.

A fundamental component of the conjectural plan for determining the generators is considering, for each prime number p , the reduction of our curve modulo p . The basic idea is quite simple: since our equation has integer coefficients, we can reduce it modulo p ³ and look for solutions in the field \mathbb{F}_p of integers modulo p . This should give a finite³ group $E(\mathbb{F}_p)$, whose structure should be easier to analyse than that of the big group $E(\mathbb{Q})$. It's a rather simple idea, but several complications⁴ arise.

The main thing that can go wrong is that the reduction modulo p may fail to be an elliptic curve. That is actually very easy to see. To tell whether the curve is elliptic (that is, if it has no singular points), one needs to look at Δ . It is perfectly possible for Δ to be nonzero (so the curve over \mathbb{Q} is elliptic) while being at the same time congruent to zero modulo p (so that the curve over \mathbb{F}_p is singular). This phenomenon is called *bad reduction*, and it is easy to come up with examples. One might take $p = 5$, and look at the curve $y^2 = x^3 - 5$. This turns out to be an elliptic curve over \mathbb{Q} , but its reduction modulo 5 is going to have a cusp. One says, then, that the curve has bad reduction at 5. In fact, the discriminant turns out to be

³It is finite because, apart from the point at infinity, there are only p^2 possible points. In fact, the maximum possible number of points is smaller than that, but that fact takes some proving.

⁴It may seem a bit perverse to dwell on the nature of these complications, but it will turn out that we need to have at least some understanding of how this goes later on.

$\Delta = -10800$, which is clearly divisible by 2, 3, and 5, so that the curve has bad reduction at each of these. (In each case, it's easy to verify that the reduced curve has a cusp.)

We want to classify the possible types of reduction, but there is one further glitch that we have to deal with before we can do so. To see what it is, consider the curve $y^2 = x^3 - 625x$. At first glance, it seems even worse than the first, and the discriminant, which turns out to be $\Delta = -15625000000$, looks *very* divisible by 5. But look what we can do: let's change variables by setting $x = 25u = 5^2u$ and $y = 125v = 5^3v$. Then our equation becomes

$$(5^3v)^2 = (5^2u)^3 - 625(5^2u),$$

which simplifies to

$$5^6v^2 = 5^6u^3 - 5^6u,$$

and hence to

$$v^2 = u^3 - u,$$

which is not only a nice elliptic curve, but has good reduction at 5. In other words, this example shows that *curves which are isomorphic over \mathbb{Q} can have very different reductions modulo p* .

It turns out that among all the possible equations for our curve, one can choose an equation that is *minimal*, in the sense that its discriminant will be divisible by fewer primes than the discriminant for other equations. Since the primes that divide the discriminant are the primes of bad reduction, a minimal equation will have reduction properties that are as good as possible. When studying the reduction properties of the curve, then, one must also pass to such a minimal equation (and there are algorithms to do this).

Well, then, suppose we have done so, and have an elliptic curve E given by a minimal equation. Then we can classify all prime numbers into three groups:

- *Primes of good reduction*: those which do not divide the discriminant of the minimal equation. The curve modulo p is an elliptic curve, and we have a group $E(\mathbb{F}_p)$.
- *Primes of multiplicative reduction*: those for which the curve modulo p has a node. If the singular point is (x_0, y_0) , it turns out that the set $E(\mathbb{F}_p) - \{(x_0, y_0)\}$ has a group structure, and is isomorphic to the multiplicative group $\mathbb{F}_p - \{0\}$.
- *Primes of additive reduction*: those for which the curve modulo p has a cusp. If the singular point is (x_0, y_0) , the set $E(\mathbb{F}_p) - \{(x_0, y_0)\}$ once again has a group structure, and is isomorphic to the additive group \mathbb{F}_p .

No curve can have good reduction everywhere, so there will always be some bad primes, but the feeling one should get is that multiplicative reduction is somehow not as bad as additive reduction. There are various technical reasons for this, which we don't really need to go into. Instead, we codify the information about the reduction types of the curve into a number, called the *conductor* of the curve. We define the conductor to be a product $N = \prod p^{n(p)}$, where

$$n(p) = \begin{cases} 0 & \text{if } E \text{ has good reduction at } p \\ 1 & \text{if } E \text{ has multiplicative reduction at } p \\ \geq 2 & \text{if } E \text{ has additive reduction at } p \end{cases}$$

(The exact value of $n(p)$ for the case of additive reduction depends on some rather

subtle properties of the reduction modulo such primes; most of the time, the exponent is 2.) The result is that one can tell, by looking at the conductor, exactly what the reduction type of E at each prime is.

The elliptic curves we will want to consider are those whose reduction properties are as good as possible. Since good reduction at all primes is not possible, we opt for the next best thing: good reduction at almost all primes, multiplicative reduction at the others. Such curves are called *semistable*:

Definition 2. *An elliptic curve is called semistable if all of its reductions are either good or multiplicative. Equivalently, a curve is semistable if its conductor is square-free.*

A crucial step in the application of Wiles' theorem to FLT will be verifying that a certain curve is semistable. Just to give us some reference points, let's look at a few examples.

1. Let E_1 be the curve $y^2 = x^3 - 5$, which we considered above. One checks that this equation is minimal, and that the curve has additive reduction at 2, 3, and 5, so that it is not semistable. The conductor turns out to be equal to 10800 (essentially, the same as the discriminant!).
2. Let E_2 be the curve $y^2 + y = x^3 + x$. This has multiplicative reduction at 7 and 13 (checking this makes a nice exercise) and good reduction at all other primes. Hence, E_2 is semistable and its conductor is 91.
3. Let E_3 be the curve $y^2 = x^3 + x^2 + 2x + 2$ (which is minimal). This has discriminant $\Delta = -1152 = -2^7 \cdot 3^2$, so that the bad primes are 2 and 3. It turns out that the reduction is multiplicative at 3 and additive at 2, and the conductor is 384; the curve is not semistable.
4. *The main example for the purpose at hand:* Let a, b , and c be relatively prime integers such that $a + b + c = 0$. Consider the curve E_{abc} whose equation⁵ is $y^2 = x(x - a)(x + b)$. Depending on what a, b , and c are, this equation may or may not be minimal, so let's make the additional assumptions that $a \equiv -1 \pmod{4}$ and that $b \equiv 0 \pmod{32}$. In this case, the equation is *not* minimal. A minimal equation for this curve turns out to be

$$y^2 + xy = x^3 + \frac{b - a - 1}{4}x^2 - \frac{ab}{16},$$

which we get by the change in variables $x \rightarrow 4x, y \rightarrow 8y + 4x$. One can then compute that the discriminant is $\Delta = a^2b^2c^2/256$ (not surprising: a constant times the product of the squares of the differences of the roots of the original cubic), and that the curve is semistable. The primes of bad reduction are those that divide abc (this would be easy to see directly from the equation, by checking when there is a multiple root modulo p), and therefore the conductor is equal to the product of the primes that divide abc :

$$N = \prod_{p|abc} p$$

⁵It may strike the reader as funny that c is absent from the equation. Keep in mind, however, that c is completely determined by a and b , so that it is really not as absent as all that. The crucial point is that the roots of the cubic on the right hand side are 0, a , and $-b$, so that the differences of the roots are (up to sign) exactly a, b , and c .

(this number is sometimes called the *radical* of abc). We will be using curves of the form E_{abc} (for very special a , b , and c) when we make the link with FLT.

We need a final bit of elliptic curve theory. It is interesting to look at the number of points in the groups $E(\mathbb{F}_p)$ as p ranges through the primes of good reduction for E . Part of the motivation for this is the reasoning that if the group $E(\mathbb{Q})$ is large (i.e. there are many rational solutions), one would expect that for many choices of the prime p many of the points in $E(\mathbb{Q})$ would survive reduction modulo p , so that the group $E(\mathbb{F}_p)$ would be large. Therefore, one would like to make some sort of conjecture that said that if the $E(\mathbb{F}_p)$ are very large for many primes p , then the group $E(\mathbb{Q})$ will be large.

Elaborating and refining this idea leads to the conjecture of Birch and Swinnerton-Dyer, which we won't get into here. But even this coarse version suggests that the variation of the size of $E(\mathbb{F}_p)$ as p runs through the primes should tell us something about the arithmetic on the curve. To “encode” this variation, we start by observing that the (projective) line over \mathbb{F}_p has exactly $p + 1$ points (the p elements of \mathbb{F}_p , plus the point at infinity). We take this as the “standard” number of points for a curve over \mathbb{F}_p , and, when we look at $E(\mathbb{F}_p)$, record how far from the standard we are. To be precise, given an elliptic curve E and a prime number p at which E has good reduction, we define a number a_p by the equation

$$\#E(\mathbb{F}_p) = p + 1 - a_p.$$

For primes of bad reduction, we extend the definition in a convenient way; it turns out that we get $a_p = \pm 1$ when the reduction is multiplicative (with a precise rule to decide which) and $a_p = 0$ when it is additive.

The usual way to “record” the sequence of the a_p is to use them to build a complex analytic function called the *L-function* of the curve E . It then is natural to conjecture that this *L-function* has properties similar to those of other *L-functions* that arise in number theory, and that one can read off properties of E from properties of its *L-function*. This is a huge story which we cannot tell in this article, but which is really very close to some of the issues which we do discuss later on. Suffice it to say, for now, that we get a function

$$L(E, s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s},$$

where the a_p are exactly the same as the ones we just introduced, the a_n are determined from the a_p by “Euler product” expansion for the *L-function*, and the series can be shown to converge when $\text{Re}(s) > 3/2$. The *L-function* is conjectured to have an analytic continuation to the whole complex plane and to satisfy a certain functional equation.

It is time to introduce the other actors in the play and to explain how they relate to elliptic curves. The reader who would like to delve further into this theory has a lot to choose from. As an informal introduction, one could look at J. Silverman's article [Sil93], which relates elliptic curves to “sums of two cubes” and Ramanujan's taxicab number. Various introductory texts are available, including [Cas91], [Hus87], [Kna92], [Sil86], and [ST92]. Each of these has particular strengths; the last is intended as an undergraduate text. In addition to these and other texts, the interested reader might enjoy looking at symbolic manipulation software that will handle elliptic curves well. Such capabilities are built into GP-PARI and SIMATH, and can be added to *Mathematica* by using Silverman's *EllipticCurveCalc* package

(which is what we used for most of the computations in this paper), and to *Maple* by using Connell's *Apecs* package. See[C⁺], [Z⁺], [SvM], [Con].

2.3 Modular Forms. Modular forms start their lives as analytic objects (or, to be more honest, as objects of group representation theory), but end up playing a very intriguing role in number theory. In this section, we will *very* briefly sketch out their definition and explain their relation to elliptic curves.

Let $\mathfrak{h} = \{x + iy \mid y > 0\}$ be the complex upper half-plane. As is well known (and, in any case, easy to check), matrices in $SL_2(\mathbb{Z})$ act on \mathfrak{h} in the following way. If $\gamma \in SL_2(\mathbb{Z})$ is the matrix

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

(so that a, b, c , and d are integers and $ad - bc = 1$), and $z \in \mathfrak{h}$, we define

$$\gamma \cdot z = \frac{az + b}{cz + d}.$$

It is easy to check that if $z \in \mathfrak{h}$ then $\gamma \cdot z \in \mathfrak{h}$, and that $\gamma_1 \cdot (\gamma_2 \cdot z) = (\gamma_1 \gamma_2) \cdot z$.

We want to consider functions on the upper half-plane which are “as invariant as possible” under this action, perhaps when restricted to a smaller group. The subgroups we will need to consider are the “congruence subgroups” which we get by adding a congruence condition to the entries of the matrix. Thus, for any positive integer N , we want to look at the group

$$\Gamma_0(N) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) \mid c \equiv 0 \pmod{N} \right\}.$$

We are now ready to begin defining modular forms. They will be functions $f: \mathfrak{h} \rightarrow \mathbb{C}$, holomorphic, which “transform well” under one of the subgroups $\Gamma_0(N)$. To be specific, we require that there exist an integer k such that

$$f\left(\frac{az + b}{cz + d}\right) = (cz + d)^k f(z).$$

Applying this formula to the special case in which the matrix is

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

shows that any such function must satisfy $f(z + 1) = f(z)$, and hence must have a Fourier expansion

$$f(z) = \sum_{n=-\infty}^{\infty} a_n q^n \quad \text{where } q = e^{2\pi iz}.$$

We require that this expression in fact only involve non-negative powers of q (and in fact we extend that requirement to a finite number of other, similar, expansions, which the experts call the “Fourier expansions at the other cusps”). A function satisfying all of these constraints is called a *modular form of weight k on $\Gamma_0(N)$* . The number N is usually called the *level* of the modular form f .

We will need to consider one special subspace of the space of modular forms of a given weight and level. Rather than having a Fourier expansion with non-negative powers only, we might require *positive* powers only (in the main expansion and in the ones “at the other cusps”). We call such modular forms *cusp forms*; they turn out to be the more interesting part of the space of modular forms.

Finally, one must make a remark on the relation between the theory at various levels: if N divides M , then every form of level N (and weight k) gives rise to (a number of) forms of level M (and the same weight). The subspace generated by all forms of level M and weight k which arise in this manner (from the various divisors of M) is called the space of *old forms* of level M . With respect to a natural inner product structure on the space of modular forms, one can then take the orthogonal complement of the space of old forms. This complement is called the space of *new forms*, which are the ones we will be most interested in.

What really makes the theory of modular forms interesting for arithmetic is the existence of a family of commuting operators on each space of modular forms, called the Hecke operators. We will not go into the definition of these operators (they are quite natural from the point of view of representation theory); for us the crucial things will be:

- For each positive integer n relatively prime to the level N , there is a Hecke operator T_n acting on the space of modular forms of fixed weight and level N .
- The Hecke operators commute with each other.
- If m and n are relatively prime, then $T_{nm} = T_n T_m$.

We will be especially interested in modular forms which are eigenvectors for the action of all the Hecke operators, i.e., forms for which there exist numbers λ_n such that $T_n(f) = \lambda_n f$ for each n which is relatively prime to the level. We will call such forms *eigenforms*.

This is all quite strange and complicated, so let's immediately point out one connection between modular forms and elliptic curves. Suppose one has a modular form which is

- of weight 2 and level N ,
- a cusp form,
- new,
- an eigenform.

If that is the case, one can normalize the form so that its Fourier expansion looks like

$$f(z) = \sum_{n=1}^{\infty} a_n q^n \quad \text{with } a_1 = 1.$$

Suppose that, once we have done the normalization,

- all of the Fourier coefficients a_n are integers.

Then there exists an elliptic curve whose equation has integer coefficients, whose conductor is N , and whose a_n are exactly the ones that appear in the Fourier expansion of f . In particular, the L -function of E can be expressed in terms of f (as a Mellin transform), and the nice analytic properties of f then allow us to prove that the L -function does have an analytic continuation and does satisfy a functional equation.

This connection between forms and elliptic curves is so powerful that it led people to investigate the matter further. The first one to suggest that *every* elliptic curve should come about in this manner was Y. Taniyama, in the mid-fifties. The suggestion only penetrated the mathematical culture much later, largely due to the work of G. Shimura, and it was made more precise by A. Weil's work pinning down the role of the conductor. We now call this the "Shimura-Taniyama-Weil

Conjecture.” Here it is:

Conjecture 1 (Shimura-Taniyama-Weil). *Let E be an elliptic curve whose equation has integer coefficients. Let N be the conductor of E , and for each n let a_n be the number appearing in the L -function of E . Then there exists a modular form of weight 2, new of level N , an eigenform under the Hecke operators, and (when normalized) with Fourier expansion equal to $\sum a_n q^n$.*

For any specific curve, it is not too hard to check that this is true. One takes E , determines the conductor and the a_n for a range of n . Since the space of modular forms of weight 2 and level N is finite-dimensional, knowing enough of the a_n must determine the form, and we can go and look if it is there. (In general, given a list of a_n , it is not at all easy to determine whether $\sum a_n q^n$ is the Fourier expansion of a modular form, so we need to do it the other way: we generate a basis of the space of modular forms, then try to find our putative form as a linear combination of the basis.) If we find a form with the right (initial chunk of) Fourier expansion, this gives *prima facie* evidence that the curve satisfies the STW conjecture. To clinch the matter, one can use a form of the Čebotarev density theorem to show that if *enough* (in an explicit sense) of the a_n are right, then they all are.

This method has been used to verify the STW conjecture for any number of specific curves (see, for example, [Cre92]). The conjecture has a really crucial role in the theory of elliptic curves; in fact, curves that satisfy the conjecture are known as “modular elliptic curves,” and many of the fundamental new results in the theory have only been proved for curves that have this property.

As our final remark on modular forms, we point out that it is possible, for any given N , to determine (essentially using the Riemann-Roch theorem) the exact dimension of the space of cusp forms of weight 2 and level N . This gives us a very good handle on what curves of that conductor should exist (if the STW conjecture is true).

For more information on modular forms, one might look at [Lan76], [Miy89], or [Shi71]. There is an intriguing account of the Shimura-Taniyama-Weil conjecture, in a very different spirit, in Mazur’s article [Maz91], and a useful survey in [Lan91].

2.4 Galois Representations. The final actors in our play are Galois representations. One starts with the Galois group of an extension of the field of rational numbers. To understand this Galois group, one can try to “represent” the elements of the group as matrices. In other words, one can try to find a vector space on which our Galois group acts, which gives a way to associate a matrix to each element of the group. This in fact gives a group homomorphism from the Galois group to a group of matrices (this need not be injective; when it is, one calls the representation “faithful”).

Rather than work with specific finite extensions of \mathbb{Q} , we work with the Galois group $G = \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ of the algebraic closure of \mathbb{Q} . This is a huge group (which one makes more manageable by giving it a topology) that hides within itself an enormous amount of arithmetic information. The representations we will be considering will be into 2×2 matrices over various fields and rings, and they will (for the most part) be obtained from elliptic curves and from modular forms.

To see how to get Galois representations from an elliptic curve, let’s start with an elliptic curve E , whose equation has coefficients in \mathbb{Z} . Choose a prime p . Since the (complex, say) points of E form a group, one can look in this group for points which are of order p (that is, for points (x, y) such that adding them to themselves

p times gives the identity). It turns out that (over \mathbb{C}) there are p^2 such points, and they form a subgroup which we denote by $E[p]$. In fact, this group is isomorphic to the product of two copies of \mathbb{F}_p :

$$E[p] \cong \mathbb{F}_p \times \mathbb{F}_p.$$

Now, the points in $E[p]$ are a priori complex, but on closer look one sees that in fact they are all defined over some extension of \mathbb{Q} , and in particular that transforming the coefficients of a point of order p by the Galois group G yields another point of order p . In fact, it's even better than that: since the rule for adding points is defined in rational terms, the whole group structure is preserved. Since $E[p]$ looks like a vector space of dimension 2 over \mathbb{F}_p , this means that each element of G acts as a linear transformation on this space, and hence that we get a representation

$$\bar{\rho}_{E,p}: G \rightarrow \mathrm{GL}_2(\mathbb{F}_p).$$

(We use a bar to remind ourselves that this is a representation “modulo p .”)

Now, $\mathrm{GL}_2(\mathbb{F}_p)$ is a finite group, and G is very infinite, so this representation, while it tells us a lot, can't be the whole story. It turns out, however, that we can use p -adic numbers to get a whole lot more. Instead of considering only the points of order p , we can consider points of order p^n for each n . This gives a whole bunch of subgroups

$$E[p] \subset E[p^2] \subset E[p^3] \subset \dots$$

and a whole bunch of representations, into $\mathrm{GL}_2(\mathbb{F}_p)$, then into $\mathrm{GL}_2(\mathbb{Z}/p^2\mathbb{Z})$, then into $\mathrm{GL}_2(\mathbb{Z}/p^3\mathbb{Z}) \dots$. Putting all of these together ends up by giving us a p -adic representation

$$\rho_{E,p}: G \rightarrow \mathrm{GL}_2(\mathbb{Q}_p)$$

which hides within itself all of the others. The representations $\rho_{E,p}$ contain a lot of arithmetic information about the curve E .

And how does it look on the modular forms side? Well, it follows from the work of several mathematicians (M. Eichler, G. Shimura, P. Deligne, and J.-P. Serre) that, whenever we have a modular form f (of any weight) which is an eigenform for the action of the Hecke operators and whose Fourier coefficients (after normalization) are integers, we can construct a representation

$$\rho_{f,p}: G \rightarrow \mathrm{GL}_2(\mathbb{Q}_p)$$

which is attached to f in a precise sense which is too technical to explain here. (The construction of the representation is quite difficult, and in fact no satisfactory expository account is yet available.)

The crucial thing to know, for our purposes, is that *when an elliptic curve E arises from a modular form f , then the representations $\rho_{E,p}$ and $\rho_{f,p}$ are the same*. In fact, a converse is also true: given a curve E , if one can find a modular form f such that $\rho_{E,p}$ is the same as $\rho_{f,p}$ then E will be modular.

3 THE PLAY. We are now ready to take the plunge and try to see how all of this theory relates to Fermat's Last Theorem. The idea is to assume that FLT is false, and then, using this assumption, to construct an elliptic curve that contradicts just about every conjecture under the sun.

3.1 Linking FLT to Elliptic Curves. So let's start by assuming FLT is false, i.e., that there exist three non-zero integers u, v , and w such that $u^p + v^p + w^p = 0$ (as we know, we only need to consider the case of prime exponent p , which is therefore odd, so that we can recast a solution in Fermat's form to be in the form above). Since we know that the theorem is true for $p = 3$, we might as well assume that $p \geq 5$. We may clearly assume that u, v , and w are relatively prime, which means that precisely one of them must be even. Let's say v is even. Since p is bigger than two, we can see, by looking at the equation modulo 4, that one of u and w must be congruent to -1 modulo 4, and the other must be congruent to 1. Let's say $u \equiv -1 \pmod{4}$.

Let's use this data to build an elliptic curve, following an idea due to G. Frey (see [Fre86], [Fre87a], [Fre87b]). We consider the curve

$$y^2 = x(x - u^p)(x + v^p).$$

This is usually known as the Frey curve. Following our discussion, above, of the curve E_{abc} , we already know quite a bit about the Frey curve. Here's a summary:

1. Since v is even and $p \geq 5$, we know that we have $v^p \equiv 0 \pmod{32}$. We also know that $u^p \equiv -1 \pmod{4}$. This puts us in the right position to use what we know about curves E_{abc} .
2. The minimal discriminant of the Frey curve is

$$\Delta = \frac{(uvw)^{2p}}{256}.$$

3. The conductor of the Frey curve is the product of all the primes dividing $u^p v^p w^p$, which is, of course, the same as the product of all the primes dividing uvw .
4. The Frey curve is semistable.

Now, as Frey observed in the mid-1980s, this curve seems much too strange to exist. For one thing, its conductor is extremely small when compared to its discriminant (because of that exponent of $2p$). For another, its Galois representations are pretty weird. Very soon, people were pointing out that there were several conjectures that would rule out the existence of Frey's curve, and therefore would prove that Fermat was correct in saying that his equation had no solutions.

3.2 FLT follows from the Shimura-Taniyama-Weil Conjecture. It was already clear to Frey that it was likely that the existence of his curve would contradict the Shimura-Taniyama-Weil conjecture, but he was unable to give a solid proof of this. A few months after Frey's work, Serre pinpointed, in a letter to J.-F. Mestre, exactly what one would need to prove to establish the link. In this letter (published as [Ser87a]), Serre describes the situation with the phrase "STW + ε implies Fermat." Because of this, the missing theorem became known, for a while, as "conjecture epsilon." This conjecture was proved by K. A. Ribet in [Rib90] about a year later, and this established the link. A survey of these results can be found in [Lan91].

What Serre noticed was that the representation modulo p

$$\bar{\rho}_{E,p}: G \rightarrow \mathrm{GL}_2(\mathbb{F}_p)$$

obtained from the Frey curve was rather strange. It looked like the sort of

representation one would get from a modular form of weight 2, but if one applied the “usual recipe” for guessing the level of that modular form, the answer came out to be $N = 2$. He also showed that the modular form must be a cusp form. The problem is that *there are no cusp forms of weight 2 and level 2!*

So suppose there is a solution of the Fermat equation for some prime p , and use this solution to build a Frey curve E . Let N be the conductor of E (which we determined above). Suppose, also, that STW holds for E , so that there exists a modular form of weight 2 and level N whose Galois representation is the same as the one for E . Then we have the following curious situation: we have a representation $\bar{\rho}$ which we know comes from a modular form of weight 2 and level N , but which *looks* as if it should come from a modular form of smaller level.

Here is where Ribet’s theorem comes in: he proves that (under certain hypotheses which will hold in our case) whenever this happens the modular form of smaller level must actually exist! Notice that this doesn’t mean that the original modular form came from lower level; what it means is that there is a form of lower level whose representation reduces modulo p to the same representation.

The upshot of Ribet’s theorem is the following:

Theorem 1 (Ribet). *Suppose STW holds for all semistable elliptic curves. Then FLT is true.*

This is true because if FLT were false, one could choose a solution of the Fermat equation and use it to construct a Frey curve, which would be a semistable elliptic curve. By STW, this curve would be attached to a modular form, so that its Galois representation is attached to a modular form. By Ribet’s theorem, there must exist a modular form of weight 2 and level 2 which gives the same representation modulo p . Just a little more work allows one to check that this modular form must be a cusp form. But this is a contradiction, because there are *no* cusp forms of weight 2 and level 2.

3.3 Deforming Galois Representations. It is now that we come to Wiles’ work. His idea was that one can attack the problem of proving STW by using the Galois representations, and in particular by thinking of “deformations” of Galois representations. The idea is to consider not only a representation modulo p , but also *all* the possible p -adic representations attached to it (one speaks of “all the possible lifts” of the representation modulo p). These can be thought of as “deformations” because, from the p -adic point of view, they are “close” to the original representation.

This sort of idea had been introduced by B. Mazur in [Maz89]. Mazur showed that one could often obtain a “universal lift,” i.e., a representation into GL_2 of a big ring such that all possible lifts were “hidden” in this representation. If one knew that the representation modulo p were modular, then one could make another big ring “containing” all the lifts which are attached to modular forms. The abstract deformation theory then provides us with a homomorphism between these two rings, and one can try to prove that this is an isomorphism. If so, it follows that all lifts are modular.

What Wiles proposes to do is very much in this spirit, except that he restricts himself to lifts that have especially nice properties. He starts with a representation modulo p , and supposes that it is modular and that it satisfies certain technical assumptions. Then he considers all possible deformations which “look like they

could be attached to forms of weight 2,” and gets a deformation ring. Considering all deformations which are attached to modular forms of weight 2 gives a second ring (which is closely related to the algebra generated by the Hecke operators, in fact). Wiles then attempts to prove, using a vast array of recent results, including ideas of Mazur, Ribet, Faltings, V. Kolyagin, and M. Flach, that these two rings are the same.

It is not hard to see that the homomorphism between the two rings we want to consider is surjective. The difficulty is to prove it is also injective. Wiles reduces this question to bounding the size of a certain cohomology group. It is here that the brilliant ideas of Kolyagin and of Flach come in. About five years ago, Kolyagin came up with a very powerful method for controlling the size of certain cohomology groups, using what he calls “Euler systems” (see [Kol91] and the survey of the method in [Maz93]). This method seems to be adaptable to any number of situations, and has been used to prove several important recent results. The initial breakthrough showing how one could begin to use Kolyagin’s method in our context is due to Flach (see [Fla92]), who found a way to construct something that can be thought of as the beginning of an Euler system applicable to our situation. Wiles called on all these ideas to construct a “geometric Euler system” which plays a central role in the argument. (*It is at this point that the current difficulty lies.*)

From the bound on the cohomology group one will get a proof that the two rings are in fact isomorphic. Translated back to the language of representations, this means that if one starts with a representation modulo p which satisfies Wiles’ technical assumptions (and is modular), then any lift of the kind Wiles considers is also modular.

3.4 Put it all together Assume, then, that one can prove that all lifts of a modular representation are still modular. Now suppose we have an elliptic curve E whose representation modulo p we can prove (by some means) to be modular. Suppose also that this representation satisfies Wiles’ technical assumptions. Then any lift of this representation is modular. But the p -adic representation $\rho_{E,p}$ attached to E is one such lift! It follows that this representation is modular, and hence that E is modular.

All we need, now, is to prime the pump: we must find a way to decide that the representation modulo p is modular, and then use that to clinch the issue. What Wiles does is quite beautiful.

First of all, he takes a semistable elliptic curve, and looks at the Galois representation modulo 3 attached to this curve. At this point, there are two possibilities. The representation, as we pointed out above, amounts to an action of the Galois group on the vector space $\mathbb{F}_3 \times \mathbb{F}_3$. Now, it may happen that there is a subspace of that vector space which is invariant under every element of the Galois group. In that case, one says that the representation is *reducible*. If not, it is *irreducible*.

One has to be just a little more careful. Just as it sometimes happens that a real matrix has complex eigenvalues, it can happen that the invariant subspace only exists after we enlarge the base field. We will say a representation is *absolutely irreducible* when this does not happen: even over bigger fields, there is no invariant subspace.

Well, look at $\bar{\rho}_{E,3}$. It may or may not be absolutely irreducible. If it is, Wiles calls upon a famous theorem of J. Tunnell, based on work of R.P. Langlands (see

[Tun81], [Lan80]) to show that it is modular, and hence, using the deformation theory, that the curve is modular.

If $\bar{\rho}_{E,3}$ is not absolutely irreducible, Wiles shows that there is another elliptic curve which has the same representation modulo 5 as our initial curve, but whose representation modulo 3 is absolutely irreducible. By the first case, it is modular. Hence, its representation modulo 5 is modular. But since this is the same as the representation modulo 5 attached to our original curve, we can apply the deformation theory for $p = 5$ to conclude that our original curve is modular.

If Wiles' strategy is successful, we get:

Theorem 2. *The Shimura-Taniyama-Weil conjecture holds for any semistable elliptic curve.*

And, since the Frey curve is semistable,

Corollary 1. *For any $n \geq 3$, there are no non-zero integer solutions to the equation $x^n + y^n = z^n$.*

Of course, this is just *one* corollary of the proof of the STW conjecture for semistable curves, and it is certain that there will be many others still. For example, as Serre pointed out in [Ser87b], one can apply Frey's ideas to many other diophantine equations that are just as hard to handle as Fermat's. These are equations that are closely related to the Fermat equation, of the form

$$x^p + y^p = Mz^p,$$

where p is a prime number and M is some integer. From Serre's argument and Wiles' result, one gets something like this:

Corollary 2. *Let p be a prime number, and let M be a power of one of the following primes:*

$$3, 5, 7, 11, 13, 17, 19, 23, 29, 53, 59.$$

Suppose that $p \geq 11$ and that p does not divide M . Then there are no nonzero integer solutions of the equation $x^p + y^p = Mz^p$.

The proof is precisely parallel to what we have done before: given a solution, construct a Frey curve, and consider the resulting modular form. Apply Ribet's theorem to lower its level, and then study the space of modular forms of that level to see if the form predicted by Ribet is there. If there is no such form, there can be no solution.

In fact, one can even get a general result, as Mazur pointed out:

Corollary 3. *Let M be a power of a prime number ℓ , and assume that ℓ is not of the form $2^n \pm 1$. Then there exists a constant C_ℓ such that the equation $x^p + y^p = Mz^p$ has no nonzero solutions for any $p \geq C_\ell$.*

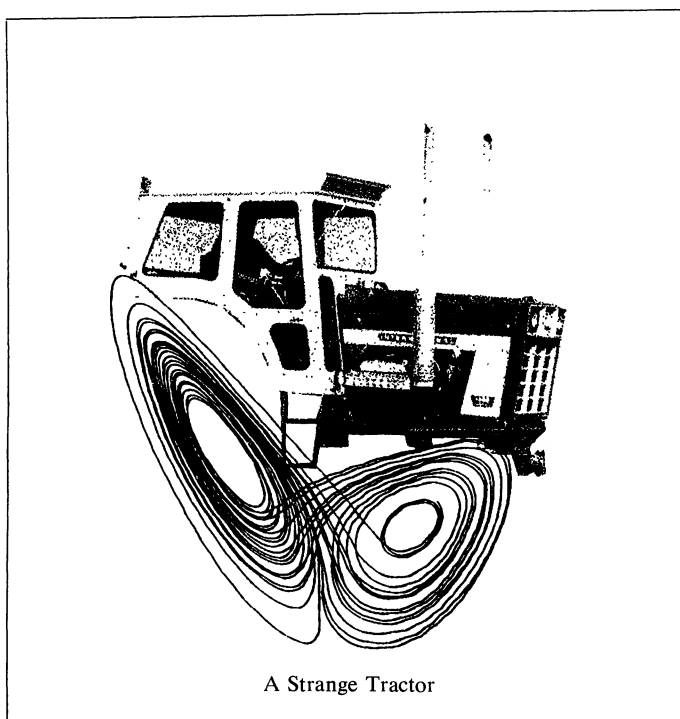
However successful they may be in the end at proving the Shimura-Taniyama-Weil conjecture, Wiles' new ideas are certain to have enormous impact.

REFERENCES

-
- [Ami75] Y. Amice, *Les nombres p -adiques*, Presses Universitaires de France, Paris, 1975.
 - [C⁺] Henri Cohen et al., *GP-PARI*, a number-theoretic “calculator” and C library. Available by anonymous ftp from math.ucla.edu.
 - [Cas86] J. W. S. Cassels, *Local fields*, Cambridge University Press, Cambridge 1986.
 - [Cas91] J. W. S. Cassels, *Lectures on elliptic curves*, Cambridge University Press, Cambridge, 1991.
 - [Con] Ian Connell, *APECS: arithmetic of plane elliptic curves*, an add-on to *Maple*. Available by anonymous ftp from math.mcgill.edu.
 - [Cre92] J. E. Cremona, *Algorithms for modular elliptic curves*, Cambridge University Press, Cambridge, 1992.
 - [CS86] Gary Cornell and Joseph H. Silverman (eds.) *Arithmetic geometry*, Springer-Verlag, Berlin, Heidelberg, New York, 1986.
 - [Fla92] M. Flach, *A finiteness theorem for the symmetric square of an elliptic curve*, Invent. Math. 109 (1992), 307–327.
 - [Fre86] G. Frey, *Links between stable elliptic curves and certain diophantine equations*, Annales Univesitatis Saraviensis, Series math. 1 (1986), 1–40.
 - [Fre87a] G. Frey, *Links between elliptic curves and solutions of $A - B = C$* , J. Indian Math. Soc. 51 (1987), 117–145.
 - [Fre87b] G. Frey, *Links between solutions of $A - B = C$ and elliptic curves*, Number Theory, Ulm 1987 (H.P. Schlickewei and E. Wirsing, eds.) Lecture Notes in Mathematics, vol. 1380, Springer-Verlag, 1987.
 - [Gou93] Fernando Q. Gouvêa, *p -adic numbers: an introduction*, Springer-Verlag, Berlin, Heidelberg, New York, 1993.
 - [Hus87] Dale Husemöller, *Elliptic curves*, Springer-Verlag, Berlin, Heidelberg, New York, 1987.
 - [Kna92] Anthony W. Knap, *Elliptic curves*, Princeton University Press, Princeton, 1992.
 - [Kob84] N. Koblitz, *p -adic numbers, p -adic analysis, and zeta-functions*, second ed., Springer-Verlag, Berlin, Heidelberg, New York, 1984.
 - [Kol91] V. Kolyagin, *Euler systems*, The Grothendieck Festschrift, vol. 2, Birkhäuser, 1991, pp. 435–483.
 - [Lan76] Serge Lang, *Introduction to modular forms*, Springer-Verlag, Berlin, Heidelberg, New York, 1976.
 - [Lan80] R.P. Langlands, *Base change for $GL(2)$* , Ann. of Math. Stud., vol. 96, Princeton University Press, Princeton, NJ, 1980.
 - [Lan91] Serge Lang, *Number theory III*, Encyclopedia of Mathematical Sciences, vol. 60, Springer-Verlag, Berlin, Heidelberg, New York, 1991.
 - [Maz89] Barry Mazur, *Deforming Galois representations*, Galois Groups Over \mathbb{Q} (Y. Ihara, K. A. Ribet, and J.-P. Serre, eds.) Springer-Verlag, 1989.
 - [Maz91] Barry Mazur, *Number theory as gadfly*, American Mathematical Monthly 98 (1991), 593–610.
 - [Maz93] Barry Mazur, *On the passage from local to global in number theory*, Bull. Amer. Math. Soc 29 (1993), 14–50.
 - [Miy89] Toshitsune Miyake, *Modular forms*, Springer-Verlag, 1989.
 - [Rib79] Paulo Ribenboim, *13 lectures on Fermat’s Theorem*, Springer-Verlag, Berlin, Heidelberg, New York, 1979.
 - [Rib90] Kenneth A. Ribet, *On modular representations of $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$ arising from modular forms*, Invent. Math. 100 (1990), 431–476.
 - [Ser87a] Jean-Pierre Serre, *Lettre à J-F Mestre*, Current Trends in Arithmetical Algebraic Geometry (Kenneth A. Ribet, ed.), Contemporary Mathematics, vol. 67, American Mathematical Society, 1987.
 - [Ser87b] Jean-Pierre Serre, *Sur les représentations modulaires de degré 2 de $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$* , Duke Math. J. 54 (1987), 179–230.
 - [Shi71] G. Shimura, *Introduction to the arithmetic theory of automorphic forms*, Princeton University Press, 1971.
 - [Sil86] Joseph H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, Berlin, Heidelberg, New York, 1986.
 - [Sil93] Joseph H. Silverman, *Taxicabs and sums of two cubes*, American Mathematical Monthly 100 (1993), no. 4, 331–340.
 - [ST92] Joseph H. Silverman and John Tate, *Rational points on elliptic curves*, Springer-Verlag, Berlin, Heidelberg, New York, 1992.

- [SvM] J. H. Silverman and P. van Mulbregt, *EllipticCurveCalc*, a *Mathematica* package. Available by anonymous ftp from gauss.math.brown.edu; contact jhs@gauss.math.brown.edu for information.
- [Tun81] J. Tunnell, *Artin's conjecture for representations of octahedral type*, Bull. Amer. Math. Soc. (N. S.) **5** (1981), 173–175.
- [Was82] Larry C. Washington, *Introduction to cyclotomic fields*, Springer-Verlag, Berlin, Heidelberg, New York, 1982.
- [Wei83] André Weil, *Number theory: an approach through history, from Hammurapi to Legendre*, Birkhäuser, 1983.
- [Z⁺] H. G. Zimmer et al., *SIMATH*, a computer algebra system with main focus on algebraic number theory. Contact simath@math.uni-sb.de for more information.

Colby College
 Department of Mathematics and Computer Science
 Waterville, ME 04901
 fqgouvea@colby.edu



*Submitted by Alberto Guzman
 Department of Mathematics
 The City College of CUNY
 New York, NY 10031*

Triangulating the Circle, at Random

David Aldous

1. INTRODUCTION. In a wonderful article [9] in this journal 38 years ago, George Pólya discussed combinatorial questions concerning triangulations of the n -gon. In particular, the number of triangulations of the n -gon is given by the $n - 1$ 'st Catalan number c_{n-1} , where

$$c_m = \frac{(2m - 2)!}{(m - 1)!m!}. \quad (1)$$

One of the interesting aspects of Pólya's paper is that it exposed readers to his newly developed theory of "figurate series". We wish to consider the idea of letting $n \rightarrow \infty$ and studying triangulations of the ∞ -gon, i.e. the circle. This question doesn't make much sense as combinatorics, but we can shift viewpoint and consider *random* triangulations of the n -gon, in which each of the c_{n-1} possible triangulations is equally likely. The purpose of this paper is to show that there exists an object "the random triangulation of a circle" which is in a natural sense the $n \rightarrow \infty$ limit of the random triangulation of the n -gon. As with Pólya [9], the exposition takes readers into some newly developed theory of the author.

Let's start by recalling a precise definition. A *triangulation* of a finite set S is a collection of nonintersecting line segments with endpoints from S such that the convex hull of S is partitioned into triangular regions. We shall be concerned only with the cases S_n consisting of the vertices of the regular n -gon inscribed in a fixed circle of unit circumference. In such a triangulation each point is linked to its neighbor on either side, and may or may not be linked to other points. For each n there is a finite set Δ_n of possible triangulation of S_n , so it makes sense to talk about a (uniform) random triangulation of S_n , where the word *uniform* emphasizes that all possible triangulations are equally likely. FIGURES 1 and 2 illustrate random triangulations for $n = 12$ and $n = 2,000$. In FIGURE 2 the printer drew a 2000-gon, but of course it looks like a circle, so it is tempting to regard FIGURE 2 as

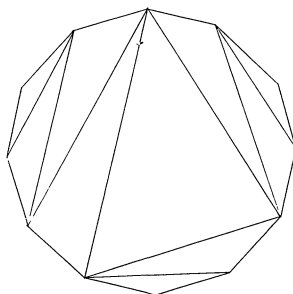


Figure 1

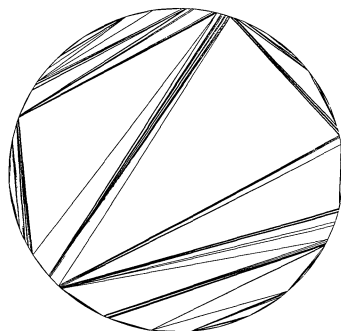


Figure 2

approximately an example of our desired limit “random triangulation of the circle”. To make sense of this object, let’s forget randomness for a while, and start by defining “triangulation of a circle”. As far as I know, the topic hasn’t been discussed before, so I get to make up my own definition.

Definition 1. *A triangulation of the circle is a closed subset of the closed disc whose complement is a disjoint union¹ of open triangles with vertices on the circumference of the circle.*

It is not hard to show that the triangulations of the circle defined above are exactly the possible limits of triangulations of n -gons. Here “limit” presupposes a topology, and we use the Hausdorff metric on compact sets:

$$d(C_1, C_2) = \sup_{x \in C_1} \inf_{y \in C_2} |y - x| + \sup_{x \in C_2} \inf_{y \in C_1} |y - x|.$$

In this “limit” assertion we regard a triangulation of the n -gon as the closed set comprised of the $2n - 3$ line segments. To illustrate, consider two sequences of triangulations of S_n . Labeling the points as 1 through n , one *possible* triangulation links

$$(1, 2)(1, 3)(1, 4)(1, 5) \dots (1, n)$$

For $n = 2,000$ (cf. FIGURE 2) the chords would look dense in the interior of the circle, and of course the $n \rightarrow \infty$ limit is the whole closed disc. Another possible triangulation, taking n to be a power of 2 for simplicity, links

$$\left(n, \frac{n}{2}\right) \left(n, \frac{n}{4}\right) \left(\frac{n}{4}, \frac{n}{2}\right) \left(\frac{n}{2}, \frac{3n}{4}\right) \left(\frac{3n}{4}, n\right) \left(n, \frac{n}{8}\right) \left(\frac{n}{8}, \frac{n}{4}\right) \left(\frac{n}{4}, \frac{3n}{8}\right) \dots$$

The $n \rightarrow \infty$ limit is a triangulation by sequential bisection of the circle, with each chord isolated and only finitely many chords longer than a prespecified length $L > 0$. But the triangulation in FIGURE 2 is qualitatively different from each of the “extreme” possibilities above: the chords are neither dense nor isolated. It turns out that the limit random triangulation of the circle, formalized as a closed subset of the closed disc, has² Hausdorff dimension $3/2$, instead of dimension 2 or 1 as in the deterministic examples above. This fact, whose proof is sketched in section 6, is

¹The union may be empty, finite or countable infinite

²With probability 1

the main concrete result of the paper. I find it remarkable that such fractal structure arises naturally³ in random combinatorial objects.

Given Definition 1, one might want immediately to pose and try to solve quantitative probability questions such as Question 1 below. Note first that the length of the longest chord in a triangulation of the circle must be at least the side-length l_0 of an inscribed equilateral triangle, and at most the diameter l_1 of the circle.

Question 1. *In a random triangulation of the circle, what is the chance that the longest chord has length greater than $(l_0 + l_1)/2$?*

This question is phrased to resemble the well-known *Bertrand's paradox*.

Question 2. *What is the chance that a random chord in the circle has length greater than l_0 ?*

This is called a paradox because, as discussed by Martin Gardner ([7] Chapter 19), three equally plausible calculations give three different answers. The conceptual point is that the notion “random chord” has no canonical meaning: instead there are several different meanings we could ascribe to it, modeling different mechanisms for physically drawing a chord in some way influenced by chance. In mathematical terms, these lead to different *probability measures* on the set of chords. The same issues arise with triangulations of the circle: before attempting problems like Question 1 we need to be clear about the probability measure underlying the word “random.” Our resolution is to use the measure which is the limit of uniform random triangulations of n -gons, and so the issue changes to proving *existence* of such a limit. This is sketched in Section 5. How to solve quantitative problems like Question 1 is discussed in Section 7.

2. CONTINUOUS FUNCTIONS AND TRIANGULATIONS OF THE CIRCLE. It turns out there is a simple way to specify a triangulation of the circle in terms of a more familiar object, viz a continuous function. Let $f: [0, 1] \rightarrow [0, \infty)$ be continuous and satisfy

$$f(0) = f(1) = 0, \quad f(t) > 0 \quad \text{for } 0 < t < 1. \quad (2)$$

Suppose t_2 is a strict local minimum of f , that is to say $f(t) > f(t_2)$ for all $t \neq t_2$ in some neighborhood of t_2 . Then by continuity there is a first time $t_3 > t_2$ at which $f(t_3) = f(t_2)$, and also a last time $t_1 < t_2$ at which $f(t_1) = f(t_2)$. Now regard $[0, 1]$ as the circumference of the circle, and draw a triangle with vertices at t_1 , t_2 , and t_3 . Repeat for each strict local minimum t'_2 . Note that if $f(t'_2) > f(t_2)$ then t'_2 is in one of the arcs (t_1, t_2) or (t_2, t_3) or (t_3, t_1) and the triangle formed by the (t'_i) lies inside the region between that arc and the corresponding edge of the triangle formed by (t_i) . This shows⁴ that triangles associated with different local minima are disjoint. So we can define a triangulation of the circle as the complement of the union of all the open triangles associated with all the local minima. Of course, if f were a polynomial we would get only a finite number of local minima and

³As opposed to fractal structures arising from recursive constructions specifically designed to produce fractals, e.g. the Cantor set

⁴More precisely, assume the values of f at different local minima are distinct, otherwise we might get both diagonals of a quadrilateral.

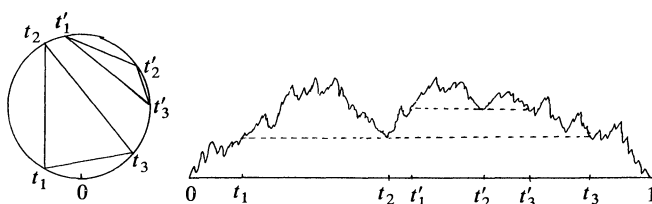


Figure 3

hence only a finite number of triangles, so the triangulation would be a closed set with non-zero area. But there exist functions f with the property

$$\{t: t \text{ is a strict local minimum of } f\} \text{ is dense in } [0, 1]. \quad (3)$$

And such functions give triangulations with zero area, which seems more natural.

This “function \rightarrow triangulation” mapping is useful for two reasons. First, it gives us a general strategy for defining random triangulations indirectly, by first defining random functions and then applying the mapping. Such an indirect approach is useful because random functions (better known as *stochastic processes*) have been the topic of half the research in mathematical probability theory for the last fifty years, so we have related a new idea to a well-studied one. Secondly, and more concretely, the mapping “function \rightarrow triangulation” turns out to be just the continuous analog of a known correspondence between triangulations of the n -gon and discrete walks, which we now describe.

3. TRIANGULATIONS, TREES, WALKS AND CATALAN NUMBERS. The combinatorial results in this section have been explained very elegantly by Martin Gardner in Chapter 20 of [8], so we shall be brief. It is convenient to consider triangulations of the $(n + 1)$ -gon, with vertex-set $S_{n+1} = \{1, 2, \dots, n + 1\}$. As mentioned before, the number of triangulations of S_{n+1} is given by the Catalan number c_n defined at (1). Various other combinatorial sets have exactly the same size, and the one of ultimate interest to us is the set W_n of positive walks of length $2n$ whose first return to 0 is at time $2n$. A *walk* has steps $+1$ or -1 : FIGURE 4 illustrates one such walk for $n = 11$. One can specify an explicit one-to-one correspondence between W_n and the set Δ_{n+1} of triangulations of S_{n+1} . This is most simply done in three stages, passing through two sets of trees whose size is also c_n . We shall specify each map in one direction only, leaving the reader to verify that it is indeed one-to-one.

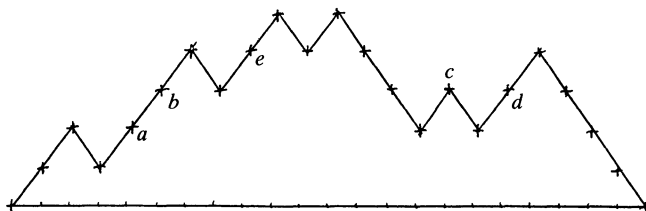


Figure 4

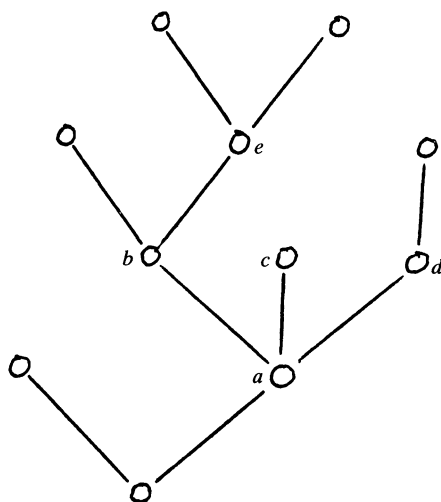


Figure 5

Map 1. This map takes the walk in FIGURE 4 to the tree in FIGURE 5. Imagine drawing a tree as the walk progresses. After the first step of the walk we draw the root of the tree. In general, when the walk makes a $+1$ step we draw a new edge from the current vertex to a new vertex. When the walk makes a -1 step we retrace our pencil from the current vertex down one edge toward the root. Thus vertices a, b, c, d, e of the tree are first drawn at the steps of the walk indicated in FIGURE 4. Note that the three children b, c, d of a are produced in a specific order as “first child, second child, third child”, and so the tree in FIGURE 5 is an *ordered tree*. Map 1 is a one-to-one correspondence between W_n and the set OT_n of rooted ordered trees on n vertices.

Map 2. FIGURE 6 shows a tree which is a *binary tree* in the following sense: each vertex is either a leaf (with no children) or an interior vertex (with exactly 2 children, distinguished as “left” and “right”). The *first child–next sibling* map takes the ordered tree in FIGURE 5 to the binary tree in FIGURE 6. Each vertex v (except the root) of the ordered tree is associated with an interior vertex v' of the binary tree. If v has children in the ordered tree then there is a first child w , and in the binary tree we make the left edge from v' go to w' ; if not, the left edge leads to a leaf. If v has a next sibling x in the ordered tree then in the binary tree we make the right edge from v' go to x' ; if not, the right edge leads to a leaf. Thus in FIGURE 6 the vertices a, b, c, d (we’ve omitted the primes) occur along a path, because b is the first child of a , then c is the next sibling of b , then d is the next sibling of c . To start the construction, the root of the binary tree is the first child of the root of the ordered tree. Map 2 is a one-to-one correspondence between OT_n and the set BT_n of binary trees with $n - 1$ internal vertices and hence with n leaves.

Map 3. There is a one-to-one correspondence between BT_n and the set Δ_{n+1} of triangulations of the points $S_{n+1} = \{1, 2, \dots, n + 1\}$. This map takes the binary tree of FIGURE 6 to the triangulation of FIGURE 1, and is illustrated by FIGURE 7 which shows the tree and the triangulation superimposed. The idea is that chords of the triangulation are identified with *edges* of the binary tree. Each chord

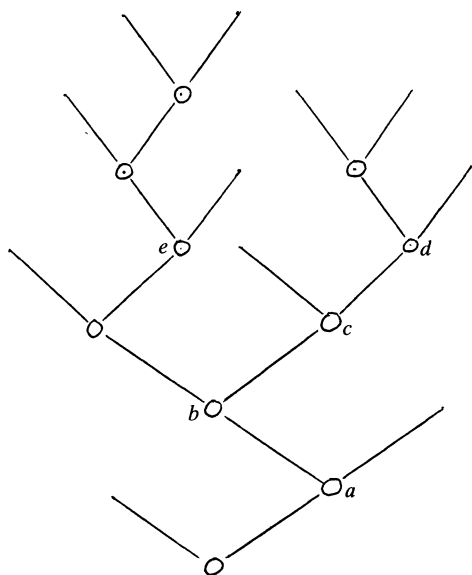


Figure 6

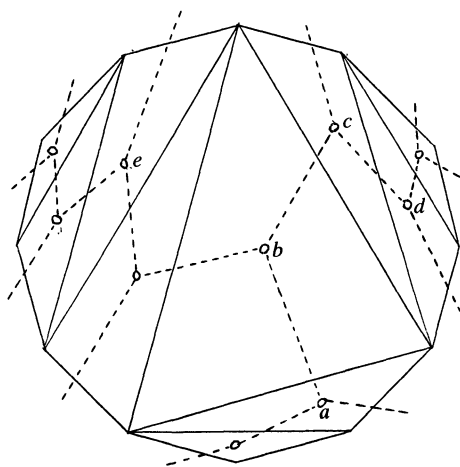


Figure 7

$(i, i + 1)$ on the boundary of the convex hull (except $(n + 1, 1)$) is identified with an edge of the binary tree leading to a leaf. Each interior vertex v of the binary tree can be identified as a point inside a triangle of the triangulation; the three edges of the tree at v correspond to the chords of the triangle, the edge leading to the root being the chord of the triangle separating the interior of the triangle from the distinguished edge $(n + 1, 1)$. Finally, the root of the tree is identified as a point in the interior of the triangle containing the edge $(n + 1, 1)$.

4. BROWNIAN MOTION AND EXCURSION. Where does this combinatorial skullduggery get us? Picking a walk at random from W_n gives us a constrained

random walk of length $m = 2n$ (constrained by positivity and the “first return at time $2n$ ” requirement). A simpler object is the corresponding unconstrained random walk of length m , where all 2^m possible walks are equally likely. Such random walks are fundamental to probability theory, and their limit *Brownian motion* is central to much of the advanced probability theory studied by mathematicians in the last fifty years. To understand the limiting area involved, consider the practical issue of drawing, for large m , the m -step walk of FIGURE 4 on a typical piece of paper with available width (left-to-right) 1 unit and height (top-to-bottom) somewhat greater. To fit the paper we clearly want to make each step have width $1/m$. It’s less clear how high up or down each step should be, but it turns out that the maximum height of the walk is of order \sqrt{m} , so a good choice is to make the steps have height $\pm \sqrt{1/m}$. For any finite m we draw a piecewise linear path, but we can now imagine the $m \rightarrow \infty$ limit as being a continuous, but “jerky” rather than “smooth”, path. This limit procedure, applied to unconstrained random walk, gives a random continuous path, called *Brownian motion*⁵. We need the variation of this result saying that the constrained random walk converges to Brownian motion constrained to satisfy (2), a process called *Brownian excursion*. Our loose description of the paths of these processes as “jerky” is reflected in precise results which say e.g. that the paths are nowhere differentiable and satisfy (3).

5. THE RANDOM TRIANGULATION OF THE CIRCLE. Granted the existence of Brownian excursion, we can see how to combine the ingredients we’ve assembled. Section 3 specifies a mapping taking constrained random walk to random triangulation of S_n . Section 2 specifies a mapping from continuous functions to triangulations of the circle, and applying this mapping to Brownian excursion gives a random triangulation of the circle. So this random triangulation of the circle must be the limit of the uniform random triangulation of S_n , because Brownian excursion is the limit of the constrained random walk.

Of course, the prose in the paragraph above skips a lot of technicalities, but the only conceptually important thing to check is that the mapping from Brownian excursion to the continuous triangulation is “essentially the same” as the mapping from the discrete walk to the discrete triangulation. To check this, consider a triangle in the triangulation of S_{n+1} for which none of the sides is very short. Using Map 3, this triangle corresponds to an interior vertex v of the binary tree. Each of the three edges from v leads toward some proportion of the leaves, and these three leaf-proportions are (give or take one leaf) the arc-lengths subtended by the sides of the triangle. In other words, a triangle of non-negligible area corresponds to a vertex v with children w, x such that, partitioning the leaves as descendants of w , descendants of x or descendants of neither, none of these three components has a negligible proportion of the leaves. Mapping⁶ now to the ordered tree using Map 2, v and x are siblings, with parent u , say, and u has a similar property in the ordered tree as v had in the binary tree. That is, if we partition the vertices of the ordered tree as descendants of one child (v) of u , or

⁵A rigorous discussion of Brownian motion and the nowhere-differentiability and dense local minima properties can be found in any good textbook on Probability at the first-year-graduate level, e.g. Durrett [6]. The Brownian excursion is less common in textbooks, but some discussion and references are in Bhattacharaya and Waymire [4].

⁶This isn’t the best way to make a rigorous argument: see the end of Section 8.1

descendants of another child (x) of u , or not descendants of u at all, then⁷ (give or take a few vertices, the descendants of the other children of u) these three proportions are the arc-lengths of the original triangle. Finally consider Map 1, taking the ordered tree to the walk. The subtrees at v and at x correspond to nearby subintervals $[\alpha_1, \alpha_2]$ and $[\beta_1, \beta_2]$ or $\{0, 1, \dots, 2n\}$ for which the walk is at the same height (1 plus the height of vertex u) at the different endpoints of the subintervals, and the walk is above that height during the subintervals. The lengths of the subintervals, relative to $2n$, are the sizes of the subtrees at v and x , and hence are approximately the arc-lengths of the original triangle. As $n \rightarrow \infty$ such subintervals become adjacent intervals $[t_1, t_2], [t_2, t_3]$ associated with a local minimum of a continuous function, and this is exactly the correspondence between functions and triangulations defined in Section 2.

6. THE FRACTAL PROPERTY OF THE RANDOM TRIANGULATION OF THE CIRCLE. Our discussion here will be very sketchy. Given $\varepsilon > 0$ consider the set S_ε of points on the circumference of the circle which are endpoints of some chord in the random triangulation with length at least ε . If we argue that S_ε has dimension $1/2$ then the reader should have no difficulty believing that the triangulation itself (a closed subset of the disc) has dimension $3/2$, because each point of S_ε corresponds to a chord in the triangulation. In terms of the construction of the triangulation from a function f chosen at random by Brownian excursion, chords correspond to intervals $[s, s']$ for which $f(s) = f(s')$ and $f(t) > f(s)$ on $s < t < s'$. (Although such an interval may not be part of a local minimum interval-pair, it will be a limit of intervals which are.) Consider such intervals straddling time 0.5: these are the intervals $[s_y, s'_y]$ where $0 < y < f(0.5)$ and

$$s_y = \sup\{t < 0.5: f(t) = y\}, \quad s'_y = \inf\{t > 0.5: f(t) = y\}.$$

So we need to argue that

$$\text{the set } \{s_y: 0 < y < f(0.5)\} \text{ has dimension } 1/2 \quad (4)$$

and then replacing 0.5 with an arbitrary rational shows that S_ε has dimension $1/2$.

Fortunately (4) can be deduced from standard facts about functions $g(t)$ chosen at random by Brownian motion. The most important of these facts is

(a) ([6] Exercise 7.4.4) The zero-set $\{t: g(t) = 0\}$ has dimension $1/2$. The most intuitive explanation of (a) is that simple random walk has order $n^{1/2}$ visits to 0 in the first n steps.

Other facts require the concept of *distribution-preserving transformation*. If a number u is chosen uniformly at random from $[0, 1]$ then the number $1 - u$ is also random and uniform on $[0, 1]$, this being a special property of the particular transformation $u \rightarrow 1 - u$. In other words, $1 - u$ has the same distribution as u . An analogous fact about Brownian motion is

(b) *Lévy's identity* ([6] Theorem 7.4.7). Given $g(t)$, define $g^*(t) = \sup_{0 \leq s \leq t} g(s)$. Then $(g^*(t) - g(t))$ and $|g(t)|$ have the same distribution. In particular, (a) and (b) imply that the set $\{t: g(t) = g^*(t)\}$ has dimension $1/2$. This

⁷The argument would break down if there were a vertex with three children, each of whose descendants comprised a non-trivial proportion of the population. But this has probability $\rightarrow 0$ as $n \rightarrow \infty$.

is essentially the same as saying

$$\text{the set } \{t_y: 0 < y\} \text{ has dimension } 1/2 \quad (5)$$

where $t_y = \inf\{t > 0: g(t) = y\}$. Next, Brownian motion has a “time-reversal” property

(c) Define $\tilde{g}(t) = g(0.5 - t)$. If $g(t)$ is chosen according to Brownian motion then so is $\tilde{g}(t)$.

This property and (5) give us (4) for functions chosen according to Brownian motion. Of course we really need (4) for Brownian excursion, but the conditioning involved in producing Brownian excursion from Brownian motion doesn’t affect “local” properties of the random functions, so (4) still holds for Brownian excursion.

7. QUANTITATIVE CALCULATIONS. Consider three vertices $1 \leq i_1 < i_2 < i_3 \leq n + 1$ of the n -gon. The chance that the triangle with these three corners occurs in the random triangulation of the $n + 1$ -gon is

$$p(n; i_1, i_2, i_3) = \frac{c_{i_2-i_1} c_{i_3-i_2} c_{n+1+i_1-i_3}}{c_n} \quad (6)$$

because the edge from i_1 to i_2 creates a (non-regular) $i_2 - i_1 + 1$ -gon outside the triangle. Noting that (1) and Stirling’s formula give

$$c_m \sim \pi^{-1/2} m^{-3/2} 2^{2m-2},$$

we can take asymptotics in (6) to get

$$p(n; i_1, i_2, i_3) \sim n^{-3} \phi(t_1, t_2, t_3) \text{ as } \left(\frac{i_1}{n+1}, \frac{i_2}{n+2}, \frac{i_3}{n+1} \right) \rightarrow (t_1, t_2, t_3)$$

where

$$\phi(t_1, t_2, t_3) = \frac{1}{4\pi} (t_2 - t_1)^{-3/2} (t_3 - t_2)^{-3/2} (1 + t_1 - t_3)^{-3/2};$$

$$0 \leq t_1 < t_2 < t_3 \leq 1. \quad (7)$$

The function ϕ represents the *frequency spectrum* of triangles in the random triangulation of the circle. That is, representing the vertices of a triangle by their distances t' around the circumference (as in Section 2),

$$\phi(t_1, t_2, t_3) dt_1 dt_2 dt_3 =$$

$$\text{mean number of triangles } (t'_1, t'_2, t'_3) \text{ with } t'_i \in [t_i, t_i + dt_i], \quad i = 1, 2, 3.$$

Various quantitative problems about the random triangulation of the circle have straightforward answers involving ϕ . Consider Question 1, and measure “length” by arc-length, so the maximal chord length is between $1/3$ and $1/2$. A moment’s thought reveals that the longest chord is just the longest edge of the triangle containing the center of the disc. So for $1/3 < x < 1/2$ the maximal chord-length is less than x iff the triangulation contains a triangle (t_1, t_2, t_3) such that

$$\max(t_2 - t_1, t_3 - t_2, 1 + t_1 - t_3) \leq x \quad (8)$$

because such a triangle necessarily contains the center of the disc. So the *probability* that the maximal chord-length is less than x is the integral of ϕ over the domain (8). With a little help from MATHEMATICA we find that the integral

has the explicit expression

$$6\pi^{-1}(\arctan 3^{-1/2} - \arctan(1 - 2x)^{1/2}) - \frac{(3x - 1)(1 - 2x)^{1/2}}{\pi x(1 - x)} \quad (9)$$

and that the probability density function of the maximal chord-length is

$$\frac{3x - 1}{\pi x^2(1 - x)^2(1 - 2x)^{1/2}}, \quad \frac{1}{3} < x < \frac{1}{2}. \quad (10)$$

The latter shows that the maximal chord-length distribution is strongly biased toward the upper end of the interval $[1/3, 1/2]$. Numerically, the median works out as 0.479 and the chance of being in the lower half of the interval (Question 1) is only 0.126.

We leave to the reader a similar problem.

Question 3. *What is the chance that the largest area of a triangle in the random triangulation is greater than half the area of a square inscribed in the circle?*

8. DISCUSSION

8.1 The weak convergence paradigm. We've discussed triangulations, but we could try the same approach in a vast range of settings. Given a sequence of discrete (typically combinatorial) random structures of size n , does there exist a continuous structure representing then $n \rightarrow \infty$ limit? If so, then for many questions about the size- n structure one can obtain the $n \rightarrow \infty$ limit by simply asking the same question of the limit structure. This is the *weak convergence paradigm*⁸. To explain the name, recall from elementary probability that the Normal distribution is, in a certain sense, the limit of Binomial distributions. This kind of convergence is called “convergence in distribution” or “weak convergence”, applied to random numbers. But we can also talk about weak convergence of random elements of an abstract metric space. This abstract theory⁹ was developed in the 1950s and 1960s, with special emphasis on the case of random functions (from $[0, \infty)$ to R) because these are just stochastic processes by another name.

Trees are fundamental in combinatorics. If you are willing to regard FIGURES 1, 4, 5 and 6 as different pictures of the same object, then we have implicitly been studying asymptotics of random trees. The weak convergence paradigm is based upon representing a “size n ” combinatorial object as an element of a metric space, scaled in such a way that the objects are comparable for different n . Informally, this is the idea of being able to draw different sized objects on the same sized piece of paper. It is trivial to do this for trees, provided you are willing to picture the tree as a triangulation or as a path (formally, the metric spaces are “closed subsets of the disc” and “continuous functions $[0, 1] \rightarrow R$ ”). But if you insist on picturing trees as in FIGURES 5 and 6, you have problems. How do you actually draw a 2000-edge tree in the style of FIGURE 5 which “looks right,” and what metric space do such trees inhabit? It's not easy¹⁰ to say. But by drawing trees as walks or triangulations we can literally *see* some interesting numerical characteristics of the

⁸My friend Mike Steele prefers to call it “the objective method”, because it involves constructing a limit object.

⁹The classic text is Billingsley [5].

¹⁰The best way I know to “draw large trees as trees” is in Aldous [1] pages 4–5.

tree, and indeed the two pictures reveal different aspects of the tree. For instance, the height of the tree is just the maximum height of the walk. And the longest chord in the triangulation corresponds to the edge of the tree which most nearly partitions the vertices into equal components.

As presented here, one might expect to get Brownian excursion asymptotics for random trees in only the two special models of random trees which arose in Section 3, viz uniform random trees in OT_n and in BT_n . In fact, much more is true. Given any model for “random tree” one can apply Map 1 to get a (non-uniform, in general) random path. But despite the non-uniformity, it turns out that the Brownian excursion limit holds for a wide class of combinatorial models of random trees. This is a remarkable fact with no simple explanation. A survey of this field, aimed at a more sophisticated audience, is in [2]. In making a rigorous argument out of the prose in Section 5, it is technically easiest to use the fact that the non-uniform random walk obtained by applying Map 1 directly to the binary tree has the Brownian excursion limit.

8.2 More about trees and triangulations. Gardner [8] mentions several other combinatorial sets whose sizes are given by the Catalan numbers, and Richard Stanley (personal communication) promises a longer list to appear in Volume 2 of [11]. Any calculation one can do with Brownian excursion leads to a limit result for random elements of any of these sets: whether such limit results are *interesting* is another matter.

Let me also mention work of Sleator et al. [10], who use the representation of binary trees as triangulations to obtain bounds on the number of “rotation steps” (a natural operation on trees as data structures in computer science) needed to move from one n -tree to another.

A rigorous treatment of some of the new results mentioned in this paper will be presented in [3].

ACKNOWLEDGMENT. My thanks to the referee for encouraging me to improve exposition and re-write the introduction (the opening paragraph was essentially provided by the referee). I also thank Persi Diaconis for helpful comments, and Jim Pitman for producing FIGURE 2.

REFERENCES

1. D. J. Aldous. The continuum random tree I. *Ann. Probab.*, 19:1–28, 1991.
2. D. J. Aldous. The continuum random tree II: an overview. In M. T. Barlow and N. H. Bingham, editors, *Stochastic Analysis*, pages 23–70. Cambridge University Press, 1991.
3. D. J. Aldous. Recursive self-similarity for random trees, random triangulations and Brownian excursion. *Ann. Probab.*, to appear.
4. R. N. Bhattacharaya and E. C. Waymire. *Stochastic Processes with Applications*. Wiley, 1990.
5. P. Billingsley. *Convergence of Probability Measures*. Wiley, 1968.
6. R. Durrett. *Probability: Theory and Examples*. Wadsworth, Pacific Grove CA, 1991.
7. M. Gardner. *Mathematical Puzzles and Diversions*. Simon and Schuster, New York, 1961.
8. M. Gardner. *Time Travel and Other Mathematical Bewilderments*. Freeman, New York, 1987.
9. G. Polya. On picture-writing. *Amer. Math. Monthly*, 51:689–697, 1956.
10. D. D. Sleator, R. E. Tarjan, and W. P. Thurston. Rotation distance, triangulations and hyperbolic geometry. *J. Amer. Math. Soc.*, 1:647–681, 1988.
11. R. P. Stanley. *Enumerative Combinatorics*. Wadsworth, Monterey CA, 1986.

Department of Statistics
University of California
Berkeley, CA 94720
aldous@stat.berkeley.edu

Hypatia and Her Mathematics

Michael A. B. Deakin

1. INTRODUCTION. The first woman mathematician of whom we have reasonably secure and detailed knowledge is Hypatia of Alexandria. Although there is a considerable amount of material available about her, very much of that is fanciful, tendentious, unreferenced or plain wrong. These limitations are to be found even in works that we might hope to be authoritative; for example, the entry in the *Dictionary of Scientific Biography* (DSB) [11]. Even where the account given is more careful and accurate [14, 19, 20], one is disappointed to be told so little of Hypatia's *Mathematics*.

This article will direct the reader's attention to the best accessible sources and will describe what is known about her mathematical activities.

2. THE HISTORICAL BACKGROUND. In about 330 B.C., Alexander the Great conquered northern Egypt and, via a deputy (Ptolemy I Soter), founded a city (Alexandria) in the Nile delta. This almost immediately became home to the Alexandrian Museum, an institution of higher learning, rather akin to the medieval universities of some 1500 years later. Euclid was an early (probably the first) "professor" of mathematics.

The Museum continued for many centuries. In 30 B.C., Cleopatra's suicide allowed the Roman Empire to occupy Alexandria, but this event destroyed neither the city's Greek heritage nor its intellectual tradition. In the years that followed, two of the greatest of late Greek mathematicians flourished in Alexandria. Diophantus was active around A.D. 250 and produced in particular his *Arithmetica* at this time. Several generations later, Pappus (c.300–c.350) also worked there.

A later mathematician, Theon of Alexandria, was the last person definitely known to have been associated with the Museum. Because he recorded two eclipses (one of the sun and one of the moon) and because he is also credited with achievements during the reign of Theodosius I, it is thought that he was at the height of his powers in the decade 360–370. Theon may well have been the last "president" of the Museum. His daughter, Hypatia, was associated with the Neo-platonic School—a different institution.

Alexandria, in the years around A.D. 400, was a turbulent mix of cultures. Christians were in the majority, but they were divided among themselves. There were also persons whom the Christians regarded as "pagans"; these could be anything from believers in the Olympian pantheon to adherents of various schools of "Neoplatonic" thought. Beyond these there were also Jews and Gnostics.

The Roman Empire, of which Alexandria was a part, was under external pressure from the Huns and the Visigoths. It split in 395 into the Western Empire (ruled from Rome) and the Eastern Empire (ruled from Constantinople). The official religion was Christianity: it had been established under Constantine. But there had been relapses; in particular, Julian the Apostate had reigned over the combined empire from 361 to 363.

At the time of Hypatia's death, the local governor was Orestes, a Christian not unsympathetic to other views, but whose authority was under challenge from that of the less tolerant Cyril (St. Cyril of Alexandria) who acceded to the bishopric in 412. The divisions that beset the city were prone to erupt into sectarian violence; the great libraries associated with the Museum were one by one destroyed, the last going up in smoke in 392 when the temple of Serapis was put to the torch during a riot. Another such disturbance was to claim Hypatia's life in the second decade of the fifth century. She died, brutally hacked to pieces, at the hands of a Christian lynch-mob.

Following this, very possibly in part because of it, the thrust of Neoplatonist thought and education moved from Alexandria to Athens. Three names require mention. Proclus (410?–485) was the last of the great mathematicians of Greek antiquity. He frequented the Neoplatonic School at Athens and is best remembered for a commentary on Book 1 of Euclid's *Elements*. After Proclus came Isidorus and his pupil Damascius (philosophers both of them rather than mathematicians, although the latter *may* have some claim on a place in mathematical history [6, pp. 312–313]). In 529, the emperor Justinian, enforcing Christianity as the state religion, closed the Neoplatonic School and Damascius went into exile in Persia.

3. THE PRIMARY SOURCES. The oldest accounts of Hypatia come to us from either the *Suda* (or *Suidae*) Lexicon or from the writings of the early Christian Church. For an accessible account of them, giving more detail than I provide here, see Mueller [14].



Medallion of Hypatia in the Introduction to Halma's edition of Theon's "Commentary on the Almageste". (Artist unknown)

Briefly, the *Suda* was a 10th-century encyclopedia, alphabetically arranged, and drawing on earlier sources. In the case of Hypatia, these are in part known. (One is a now lost work, a life of Isidorus by Damascius.) The relevant entry is unusually long, but is not seen as reliable in all its aspects (see [25]); indeed in places it contradicts itself.

"The other sources are to be found in the main in a compilation known as the *Patrologiae Graecae* [13], or PG for short. This gives earlier accounts (particularly of her death) than are available in the *Suda* and also preserves letters to her and about her from the hand of one of her pupils, Synesius of Cyrene. Also by Synesius is a letter published as a separate document included with the others in FitzGerald's translation [4].

4. LIFE AND LEGEND. The best-recorded event in Hypatia's life is her death and the manner of it. The fullest account tells us that a crowd of Christian zealots led by one Peter the Reader seized her, stripped her and proceeded to dismember

her and burn the pieces of her corpse. Another says she was burned alive, but this would seem to be a less accurate version.

The political background to this action has been the cause of much speculation. Gibbon [5] is by no means alone in attributing the guilt for the murder to Cyril, but Rist [20] disputes this, which does mean taking issue with the *Suda*. Rist's account, in essence, has it that, like victims of violence in Belfast or Beirut today, she was seized not with any great selectivity at all, but rather because she was a well-known public figure, prominent on the other side of a religious divide. This to my mind is quite compatible with the statement quoted by Gibbon to the effect that she was killed because of her outstanding ability. We need not posit any specific jealousy to say this, and Rist thinks it is unlikely that precise differences of doctrine led to her death. Rist does toy with the idea that her mathematical activities were a partial cause, hypothesizing that these included astrology. This, to me, sets us on a path we have no reason to travel.

The *date* of her death is now generally accepted to have been 415, although others have been suggested. See Mueller [14] for details.

The date of her birth is much less certain. (This is to be expected—people are not, generally speaking, famous when they are born.) The eclipses described by Theon, Hypatia's father, have been dated to 364. So, from the eclipses to the time of her death is an interval of 51 years. Valesius, an early commentator on the PG who had the wrong date for the eclipses, reckoned this interval at 47 years; rounding this to 45 produces a date of c.370, which is the generally-stated figure. Of course, astrology aside, we have no real reason to suppose that her birth coincided with the eclipses; nor have we any idea how old Theon (or more importantly his wife) was in 364. (I tend to agree with Mueller that a date of c.350 is more plausible.)

As to her life between these uncertain dates, we may readily summarize. She was a respected and eminent teacher, charismatic even, and beloved of her pupils (e.g., Synesius). We have evidence that she was regarded as physically beautiful, that she wore distinctive academic garb, that she taught not only mathematics but



also Philosophy, that she gave public lectures and may have held some kind of public office.

She seems to have been determinedly celibate, indeed repelling one ardent suitor by confronting him with one of her used menstrual pads and lecturing him on the shameful and unclean nature of what he thought beautiful (the vagina).

Although almost all the primary sources are Christian and tell of the life and death (at Christian hands) of a prominent advocate of a rival philosophy, they do so in such a way that we are left with a favorable impression of her. My reading of this is that the official discouragement of her teachings on the part of the Church authorities and of their (Christian) civic counterparts was far from complete.

Certainly that favorable impression has informed various works of literature of which the best-known in English are Kingsley's novel [10] and the passage from Gibbon. Also fiction is Hubbard's telling of Hypatia's story [9]. It formed a chapter in a popular reader early this century and has given us the most widely disseminated "portrait" of Hypatia, attributed to an artist called Gasparo, of whom I am able to learn nothing. (Of course such "portraits" have exactly the same validity as (e.g.) Doré's illustrations of the Bible.)

5. HYPATIA'S PHILOSOPHY. The Philosophy expounded by Hypatia is known to have been Neoplatonist. There were various versions of Neoplatonism, all endowing Plato's Theory of Forms with an explicitly religious dimension. Richeson [19] describes one such system; Rist [20] suggests that Hypatia actually preached another.

Richeson does however make a particularly insightful remark on the connection between Neoplatonist Philosophy and Mathematics. The nature of Mathematics is to abstract—to derive *ideas* from material things. Thus Geometry, although it has its *origin* in the practical world of land surveyors and inspectors of weights and measures, transcends these beginnings. The *Elements* deals with a world that is no longer the world of the practical but rather the world of ideas. Thus Mathematics could be seen as a paradigm of that transcendence over the material that Neoplatonism enjoined.

6. HYPATIA'S MATHEMATICS. That Hypatia was a mathematician is beyond doubt. The PG tell us that she learned her Mathematics from her father Theon and went on to excel him in the subject and to teach it to numerous students. Another such source is more critical: "Isidorus greatly outshone Hypatia, not just because he was a man and she a woman, but in the way a genuine philosopher will over a mere geometer." This opinion, which will earn no praise from either women or mathematicians, is thought to derive from Damascius' life of Isidorus, the lost work that in part informed the *Suda*. (Marrou [12], following Tannery [25], supplies the following delightful gloss: "[it] means in plain language that Isidorus knew nothing of mathematics.")

However, the *Suda* itself gives the most explicit account of Hypatia's mathematical works. It attributes to her the authorship of three works. The only things she is known to have written all deal with Mathematics or Astronomy. The books that many feel she must have authored on Neoplatonist Philosophy receive no mention. Others (e.g., Kramer [11]) have credited her with further works of Mathematics. For this there is no evidence, except in one specific instance to be described below. The relevant passage in the *Suda* is precisely twelve words long. And even this short excerpt is the subject of various alternative and disputed readings. However, there is a general consensus that Tannery [25] is correct in rendering it thus: "She

wrote a Commentary on Diophantus, [one on] the astronomical Canon, and a Commentary on Apollonius's *Conics*."

"Commentaries" were what we would now refer to as "Editions" (with the obvious difference that they needed to be copied by hand), and the author of a "Commentary" is perhaps better referred to as an "Editor." Such "Editors" or "Commentators" did, however (to a greater or lesser extent, and with greater or lesser care to distinguish their own contributions from the original), provide new material of various sorts (witness Fermat's famous marginal note to Diophantus). It should be noted that in many cases the original text has come down to us only through Commentaries or translations (often into Arabic).

Theon, Hypatia's father, was a prolific author of Commentaries. He wrote one on the *Elements* (which, in places, still provides our present text), on two other works by Euclid, the *Data* and the *Optics*, and on two works by Ptolemy, the *Almagest* and the *Handy Tables*. There were also works now lost or partly so; particularly germane to our story is a work on the astrolabe. For this and more, see Toomer [28].

The picture that emerges of Theon is one of an editor, teacher and textbook-writer rather than a research mathematician. So is he judged, often with more than a hint of disapproval. But this should not mean that his was a wasted life. His works were preserved, presumably because they were perceived as having lasting value. It is all too understandable, given the politics of late 4th-century Alexandria and the decay of the Museum, that the emphasis on research (possible in Pappus's time) should be replaced by the priority of conserving knowledge.

After considering her works *seriatim*, I shall offer the hypothesis that in her scholarly priorities Hypatia was very much her father's daughter. This, as I hope I have just made clear, is not to denigrate her.

7. APOLLONIUS'S CONICS. Apollonius lived around 200 B.C. and the *Conics* is the most important of his surviving works. See, for more detail, Toomer's account [27]. There are very few sources for our present text and Hypatia's Commentary is not one of them. Of the eight books that make up the *Conics*, the first four survive via a Commentary by Eutocius while three of the remaining four have come down to us via the Arabic. The other is lost, as is also, we must conclude, Hypatia's Commentary, unless it is the lost original of Eutocius' work.

8. THE ASTRONOMICAL CANON. In the case of the "Astronomical Canon," we are much better placed. It is now generally assumed that Tannery's interpolation (the words in brackets in Section 6) in the *Suda* entry is correct. This means that this work also was a Commentary. The most likely original is one of the works of Ptolemy, either the *Almagest* or the *Handy Tables*. It will be remembered that Theon wrote commentaries on both these works.

Theon's commentary on the *Almagest* has been printed in various editions. The best and most recent is by A. Rome [21, 22]. (But see also [23].) It comprises separate Commentaries on the thirteen books that go to make up the *Almagest*. The titular inscriptions (as described by Rome from his study of the manuscripts) of the first and second books ascribe these works to Theon himself. Books 4–13 contain no inscriptions. Only the very best manuscripts contain the Commentary on Book Three, and here the inscription tells us that the work is Theon's "in the recension of my philosopher-daughter Hypatia."

Heath [8], reviewing Rome's work, thus ascribed this chapter of the Commentary to Hypatia, with the inference that it was also the work alluded to in the *Suda*,

and that Theon (recognizing his daughter's work as superior to his own) had suppressed his earlier effort in favor of hers. (The pity, from our point of view, is that we don't have both versions before us; so we cannot see for ourselves where or how or to what extent Hypatia's Commentary differed from Theon's.) Rome himself discusses the matter at considerable length in his later work [22], but in such a way as not to rule out a possibility that has been canvassed: that father and daughter collaborated.

Neugebauer [16, p. 838] accepts this as likely. However, he regards it as probable that what the *Suda* refers to is a commentary not on the *Almagest* at all, but on the *Handy Tables*. This is because the same word (*Canon*) is used for both works. (Delambre [2] had earlier noted this same concordance of wording, but as his work predates Tannery's suggested interpolation, he credits Hypatia with a set of Astronomical Tables.) If the *Suda* were referring to a Commentary on the *Almagest*, so the argument goes, then it would speak of the *Syntaxis*, rather than the *Canon*. (*Syntaxis* is the Greek name for the work we now know by its Arabic designation.) Against this, however, is the Canon of Parsimony and the fact that Book 3 of the *Almagest* has a strongly tabular character.

9. DIOPHANTUS' ARITHMETIC. We may also have some of Hypatia's own writing from the Commentary on Diophantus. Diophantus' major work is the *Arithmetic*, originally comprising thirteen books. Of these only six now survive from the Greek, and possibly part of another, now listed as separate, the *Polygonal Numbers*. Tannery [26] suggested that all existing manuscripts known to him derived from a common source and that that source was Hypatia's Commentary. His careful "family tree" of the manuscripts was later modified in one detail and made available in the amended form in Heath's Edition [7]. The presumption was that Books 7–13 are lost because Hypatia's Commentary did not include them, much as Eutocius' Commentary extended only to the first four books of the *Conics*. This hypothesis enjoyed a deal of support, and Vogel's Article on Diophantus in the DSB simply accepts it.

The basis for this theory was the Greek text and the fact that the *Suda* reference to Hypatia's Commentary is the only mention of so ancient an edition. Sesiano [24, pp. 71–75] however queries this account. This is a matter of great controversy. The old theory will be presented first, but see the remarks at the end of this section.

On the old story, the mathematical world of today owes Hypatia a great debt, for without her we would have much less of the works of Diophantus. But there is an obvious corollary. If what survives for us is Hypatia's Commentary, then some of her work may appear there. It may be possible to see what is hers. One complication is that a later scribe was thought to have attempted to reconstruct Diophantus' original text and thus to have systematically omitted material he judged to be interpolated. But "the distinction between text and scholia being sometimes difficult to draw, he included a good deal which should have been left out" [7, p. 14].

On this account, the most likely of the supposed interpolations to have come from Hypatia's hand are two "student exercises" at the start of Book II. The first asks for the solution of the pair of simultaneous equations:

$$x - y = a, \quad x^2 - y^2 = (x - y) + b,$$

where a, b are known. The next is a minor generalization. It requires the solution

of the pair of simultaneous equations:

$$x - y = a, \quad x^2 - y^2 = m(x - y) + b,$$

where a , m and b are known. There is some evidence to link this problem to Hypatia: a nine-word phrase in the original Greek is identical with one from Euclid's *Data*, which her father had edited.

Recent work by Roshdi Rashed, Sesiano and others has suggested that some of the lost books of Diophantus in fact survive in Arabic translations. This has led to very great and indeed bitter controversy. What is at issue (apart from the personal rivalries involved) is whether Diophantus or someone else wrote the newly discovered works and where they might fit into the fragment previously published. Sesiano and others are inclined to the view that if anything of Hypatia's Commentary survives then it survives in the Arabic. There are no clear indications of what might be by her and what by Diophantus or by other scholiasts. Many of Sesiano's conclusions are hotly disputed by Rashed [18]. However, tentative attributions of material to Hypatia all tend to accept the overall assessment reached above—that her contributions to mathematical knowledge itself were slight or non-existent.

10. THE ASTROLABE. The other source for information about Hypatia's mathematical activities is the correspondence of Synesius.

There is a brief but telling reference to Hypatia in Synesius' essay-letter *De Dono Astrolabii*. The name "astrolabe" was a term applied to a variety of instruments. For a good overview of later developments, see [17]; earlier ones are discussed by Neugebauer [15]. A simple attempt to replicate the motions of the heavens in a mechanical model produces the device known as an "armillary sphere". Such an object is necessarily 3-dimensional and unwieldy, more suitable for display purposes than for use as a practical instrument of observation or computation.

However, once we have a theory of stereographic projection, the way is open for the construction of a more practical two-dimensional device. This theory was given by Ptolemy in his *Planisphaerium*, which even includes tabular material. Whether Ptolemy went on to develop the "little astrolabe" (i.e. the practical instrument) has been argued. Neugebauer regards it as probable that he did.

The next figure is Theon. Ptolemy died in about 170 A.D., about two centuries before Theon's active period. Theon wrote, as we have seen, Commentaries on the *Almagest* and the *Handy Tables*. The *Suda* also credits him with a treatise on the little astrolabe, and Arab sources refer in addition to a work of his on the armillary sphere. This set corresponds *exactly* to the set of works assigned to Ptolemy by the Arabs.

There is thus considerable evidence that Theon was familiar with the theory of the little astrolabe. We might speculate that he invented it, but the picture of Theon that has come down to us is one of Theon as a disseminator and conserver of knowledge, rather than an innovator. Moreover, Neugebauer has given us grounds to believe Ptolemy to have been the inventor.

Although Theon's work on the astrolabe is now regarded as lost, Neugebauer finds such similarities between later works that they must derive from a common source. This source he believes to be Theon. He further argues (because of the exact correspondence described above) that what Theon wrote was a Commentary on an earlier book by Ptolemy.

This gives us the background to Synesius' *De Dono Astrolabii*. Writing to Paionos, he states that he designed the astrolabe himself with help from Hypatia

and had it crafted by the very best of silversmiths. The inference is that the theory of the astrolabe and the details of its construction were passed down from Ptolemy, via Theon, to Hypatia, who in her turn taught Synesius.

11. THE HYDROSCOPE. Letter 15 of Synesius begins: “I am reduced to this, that I have to have a hydroscope.” The letter then goes on to ask her to make him one, to quite detailed specifications. The question of what he needed is puzzling. The general presumption is that he was ill.

The term “hydroscope” usually implies a *clepsydra* or water-clock, but this seems inappropriate as a translation in this case. Why should he be, even if brought so low, in such urgent need of a water-clock? FitzGerald believes that Fermat (yes, *the* Fermat) [3] is right in suggesting that what Synesius needed was a hydrometer, that is to say, a densimeter. This makes much more sense of the specifications, which refer to the need to measure the *weight* of the water (the *clepsydra* measures the *volume*), and describe an instrument that sounds very like a hydrometer.

The suggestion is that Synesius needed it in his illness somehow to measure a medicine he was taking (or less plausibly the salinity of his drinking water). Hydrometers are now used, as they well may then have been, to measure the alcoholic contents of fermented or distilled liquors. Possibly Synesius was making his own medicine by some such means. My friend and colleague Charles Hunter (Department of Anatomy, Monash University) however offers a novel suggestion—that the “hydroscope” was in fact a urinometer and that the dosage of some diuretic was calculated by reference to the specific gravity of the urine.

12. ASSESSMENT. What we know of Hypatia is little enough; what we know of her Mathematics is only a small subset of that little. There is evidence that she was greatly regarded as a teacher and a scholar. The range of her acknowledged expertise was considerable. She edited works of Geometry, Algebra and Astronomy, knew how to make astrolabes and “hydrosopes”, and did a lot else besides. One cannot but be impressed with this breadth of interest. Moreover, at the time of her death (assuming with Toomer [28] that Theon pre-deceased her) she was in fact the greatest mathematician then living in the Greco-Roman world, very likely the world as a whole.

She is variously described as a philosopher, a teacher of Philosophy, a mathematician and astronomer, a learned woman and a geometer.

We can understand the term “philosopher” in two senses: it has the technical sense that it retains to this day, but it also has a generic meaning of “thinker”. Theon also is described in the sources as a philosopher. But this is surely in the second sense; Theon clearly emerges as a specialist mathematician and astronomer—the *Suda* goes on to say as much. Hypatia does not (unless one accords weight to the quote in Section 6 above); the *Suda* is at some pains to make this clear. “She also took up other [non-mathematical] branches of philosophy and though a woman she cast [an academic robe] around herself and appeared in the centre of the city” (Rist’s translation)—the *Suda* then proceeds to describe the Philosophy she taught, mentioning the work of Plato and Aristotle, in particular.

However, if we restrict consideration to Mathematics alone, we may well query the usual judgment that Hypatia outclassed her father. It comes from the PG and modern sources regularly repeat it uncritically. We may also deduce it from Theon’s heading to his Commentary of Book 3 of the *Almagest*.

We may still however dispute this opinion and indeed argue the opposite. That a fond father might recognise and promote his daughter's improvement of one of his own works is understandable enough. That ecclesiastical historians, of whom we have no evidence of mathematical ability, might use fame or even notoriety as an index of talent is equally so. But this does not end the matter.

While it is of course too much to posit a universal theory of natural selection of scholarly works (it being by no means *always* true that the best works are the survivors) nonetheless scholars of earlier times preserved, translated and taught from those works they adjudged as valuable. Much as we do today. In fact, we do know something of the principle of natural selection that operated. Because the focus had moved from research to conservation, those works were preserved that were well regarded as *textbooks* [29]. Many research works from the period are lost.

We have no evidence of research Mathematics on the part of either father or daughter. What we can reconstruct of their Mathematics suggests to us that they edited, preserved, taught from and supplied minor addenda to the works of others. A great deal of Theon's work survives and at most a small part of Hypatia's. In other words Theon was seen as the better text-writer, even if he himself generously demurred in one case.

Where Hypatia *does* quite clearly outshine Theon is in her reputation as a teacher. She was revered as such and no similar endorsement of Theon has come down to us. (It is perfectly possible that this is the basis of the original statement.) We are left with a well-attested account of a popular, charismatic and versatile teacher. And that, I suggest, is the best picture we can form of her.

ACKNOWLEDGMENT. The present article is an abridgement of a longer original [1]. I am grateful to G. J. Tee of the University of Auckland for his careful criticism of the earlier draft and for bringing reference [14] to my attention.

REFERENCES

1. Deakin, M. A. B., "Hypatia the Mathematician", *Monash University History of Mathematics Pamphlet* 52 (1991).
2. Delambre, J. B. J., *Histoire de l'Astronomie Ancienne* 2 (New York: Johnson, 1965; reprint of an 1817 original), 317.
3. Fermat, P., *Œuvres* 1 (Ed. P. Tannery and C. Henry) (Paris: Gauthier, 1891), 352–365.
4. FitzGerald, A. (ed. and trans.), *The Letters of Synesius of Cyrene* (London: Oxford University Press, 1926).
5. Gibbon, E., *The Decline and Fall of the Roman Empire* (first published 1776–1788; many subsequent editions), Chapter 47.
6. Gow, J., *A Short History of Greek Mathematics* (New York: Chelsea, 1968; reprint of an 1884 original), 312–313.
7. Heath, T. L., *Diophantus of Alexandria* (Cambridge University Press, 1885; Dover reprint, 1964).
8. Heath T. L., Review of Ref. [21], *Class. Rev.* 52 (1938), 40.
9. Hubbafd, E., *Little Journeys [to the Homes of Great Teachers]* 23 (4) (East Aurora, NY: Roycrofters, 1908).
10. Kingsley, C., *Hypatia* (first published 1851; many subsequent editions).
11. Kramer, E. E., Article on Hypatia, *DSB* 6, 615–616.
12. Marrou, H. I., "Synesius of Cyrene and Alexandrian Neoplatonism", in *The Conflict between Paganism and Christianity in the Fourth Century* (Ed. A. Momigliano) (Oxford: Clarendon, 1963), 126–150.
13. Migne, J.-P. (ed.) *Patrologiae Graecae* (Paris: Migne, 1857–1866).
14. Mueller, I., "Hypatia", in *Women of Mathematics: A Biobibliographic Sourcebook* (Ed. L. S. Grinstein and P. J. Campbell) (New York: Greenwood, 1987).
15. Neugebauer, O., "The Early History of the Astrolabe", *Isis* 40 (1949), 240–256.

16. Neugebauer, O., *A History of Ancient Mathematical Astronomy* (Berlin: Springer, 1975).
17. North, J. D., "The Astrolabe", *Sci. Am.* 230 (1), (Jan. 1974), 96–106.
18. Rashed, R., Review of Ref. [24], *Math. Rev.* 85h:01006.
19. Richeson, A. W., "Hypatia of Alexandria", *Nat. Math. Mag.* 15 (1940), 74–82.
20. Rist, J. M., "Hypatia", *Phoenix* 19 (1965), 214–225.
21. Rome, A., *Commentaires de Pappus et de Théon d'Alexandrie sur l'Almageste* (2), Studi e Testi (Vatican) 72 (1936).
22. Rome, A., *Commentaires de Pappus et de Théon d'Alexandrie sur l'Almageste* (3), Studi e Testi (Vatican) 106 (1943).
23. Rome, A., "Le troisième livre des commentaires sur l'Almageste par Théon et Hypatie", (Paris: Presses universitaires de France, 1926; excerpted from *Ann. Soc. Sci. Bruxelles* 46, 1926)
24. Sesiano, J., *Books IV to VII of Diophantus' Arithmetica* (New York: Springer, 1982).
25. Tannery, P., "L'article de Suidas sur Hypatia", *Ann. Fac. Lettres Bordeaux* 2 (1880), 197–200.
26. Tannery, P., *Diophanti Alexandrini opera omnia* (Leipzig: Teubner, 1893–1895).
27. Toomer, G. J., Article on Apollonius, *DSB* 1, 179–193.
28. Toomer, G. J., Article on Theon, *DSB* 13, 321–325.
29. Toomer, G. J., "Lost Greek Mathematical Works in Arabic Translation", *Math. Intell.* 6 (2) (1984), 32–38.

Addendum: Too late for mention in the main article, I was made aware of the lengthy discussion of Hypatia by W. R. Knorr [*Textual Studies in Ancient and Medieval Geometry* (Boston: Birkhauser, 1989)]. Beginning from a stylistic analysis of Book Three of Theon's Commentary on the *Almagest*, Knorr builds an elaborate and detailed, though speculative, argument to attribute several other lost works to Hypatia. In particular, he suggests that Eutocius' Commentary on Apollonius' *Conics* in fact derives from Hypatia's earlier Commentary, the one mentioned in the *Suda*.

I thank Win Frost of the University of Newcastle (Australia) for bringing Knorr's work to my attention.

Department of Mathematics
Monash University
Clayton, Vic. 3168
Australia

"The main duty of the historian of mathematics, as well as his fondest privilege, is to explain the humanity of mathematics, to illustrate its greatness, beauty and dignity, and to describe how the incessant efforts and accumulated genius of many generations have built up that magnificent monument, the object of our most legitimate pride as men, and of our wonder, humility and thankfulness, as individuals. The study of the history of mathematics will not make better mathematicians but gentler ones, it will enrich their minds, mellow their hearts, and bring out their finer qualities."

—G. Sarton

Calculus II and Euler also (with a Nod to Series Integral Remainder Bounds)

Richard Barshinger

A classical theorem of Maclaurin and Cauchy [8, p. 45] states that, if $f(x)$ is positive and decreases to zero, then

$$\gamma_f = \lim_{n \rightarrow \infty} \left\{ \sum_{k=1}^n f(k) - \int_1^n f(x) dx \right\}$$

exists. The constant γ_f is called an Euler constant, the original belonging to $f(n) = 1/n$ and having the value $\gamma = 0.57721566490\dots$ [γ arises, in applied mathematics, in the formulation of Bessel functions and the gamma function, among others.] Papers such as [2], [3], [4], and [14] have discussed these constants, as well as more generally considering rates of growth, for divergent series in particular.

This paper considers a suitable approach by which the computation of the canonical Euler constant (corresponding to $\lim_{n \rightarrow \infty} \{\sum_{k=1}^n 1/k - \ln n\}$, a classical indeterminacy of the form $\infty - \infty$) can find its way, in a pedagogically sound and useful fashion, into a first calculus course. Though a number of authors have used various intermediate to advanced techniques to explore some aspects of $\sum_{n=1}^{\infty} 1/n$ and γ , we use an elementary geometrical technique for accelerating the convergence of the above limit.

Evaluation of γ , Method O (brute force—almost!). Any student knows—or has to be taught—that most any computational device will claim that an infinite series whose terms decrease to zero will converge. Consequently, it is not sufficient simply to compute successive evaluations of $\sum_{k=1}^n 1/k - \ln n$. It is, however, clear that, for all n ,

$$\sum_{k=1}^n \frac{1}{k} - \ln(n+1) < \sum_{k=1}^n \frac{1}{k} - \ln n.$$

It is easy to show that the two sequences are strictly monotonically increasing and monotonically decreasing, respectively (see, for example, [13], p. 669), and that both converge to γ . The following table of bounds can easily be classroom generated, when using, for example, the Sequences program in MicroCalc [6] and run on a 33 Mhz 486 microcomputer.

n	γ			
1800	0.576938	0.577493	0.577	(rounded)
2313	0.577000	0.577423	0.577	(truncated)
7557	0.577150	0.577282		
14778	0.577182	0.577249	0.5772	(rounded)

Computing about 15,000 terms of each of two sequences is, if nothing else, rather esthetically unattractive, however; so, alternatively, we might average the above pairs of values, which gives $\gamma = 0.577$ (truncated) as early as the 10th iterate. [This

is based on $\sum_{k=1}^n 1/k - \ln \sqrt{n(n+1)}$ and involves the geometric mean between n and $n+1$.]

Error Bounds for γ , Method 1. From a problem in [12, p. 344 & p. 621], based on FIGURE 1, it easily follows that $1/2 < \gamma < 1$. It is quite straightforward to improve substantially on these very crude bounds. FIGURE 2 shows that we may calculate a lower bound for γ by drawing secant lines from the endpoints of $f(x) = 1/x$ on the interval $[n, n+1]$ to the midpoint of the curve. The area above the secant lines and below the horizontal line $y = 1/n$ is smaller than the contribution to the value of γ by taking the area between the curve and $y = 1/n$. Similarly, an upper bound for γ may be obtained by calculating the area between $y = 1/n$ and the two intersecting lines that are drawn tangent to the curve, at $x = n$ and $x = n+1$ respectively. Summing over n , we obtain

$$\frac{1}{4} \sum_{n=1}^{\infty} \left[\frac{1}{n} - \frac{1}{n+1} \right] + \frac{1}{2} \sum_{n=1}^{\infty} \frac{1}{n(2n+1)} < \gamma < \sum_{n=1}^{\infty} \frac{1}{n(2n+1)}. \quad (1)$$

Papers such as [1], [5], [9], [10] & [11] have considered various ways to accelerate the convergence of a series. Let $f(n)$ be the generator of the terms of a convergent series and R_n be the remainder after the n th partial sum S_n . Assume f and $|f'|$ both decrease to zero (f is concave up). These inequalities, presented for archival purposes, then summarize the papers cited above.

$$\begin{aligned} \int_{n+1}^{\infty} f &< \int_n^{\infty} f - \frac{1}{2}f(n) < \int_{n+1}^{\infty} f + \frac{1}{2}f(n+1) < \int_{n+1}^{\infty} f + \frac{3}{4}f(n+1) - \frac{1}{2} \int_{n+1}^{n+3/2} f \\ &< R_n < \int_{n+1/2}^{\infty} f + \frac{1}{4}f(n+1) - \frac{1}{2} \int_{n+1/2}^{n+1} f \\ &< \int_{n+1}^{\infty} f + \frac{1}{2}f(n+1) + \frac{1}{8}|f'(n+1)| \\ &< \int_{n+1/2}^{\infty} f < \int_n^{\infty} f - \frac{1}{2}f(n+1) - \frac{1}{4}|f'(n+1)| < \int_n^{\infty} f - \frac{1}{2}f(n+1) < \int_n^{\infty} f \end{aligned} \quad (2)$$

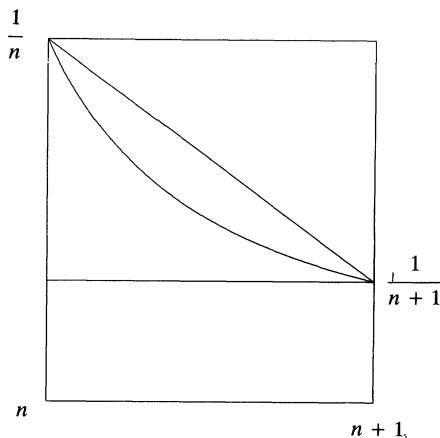


Figure 1

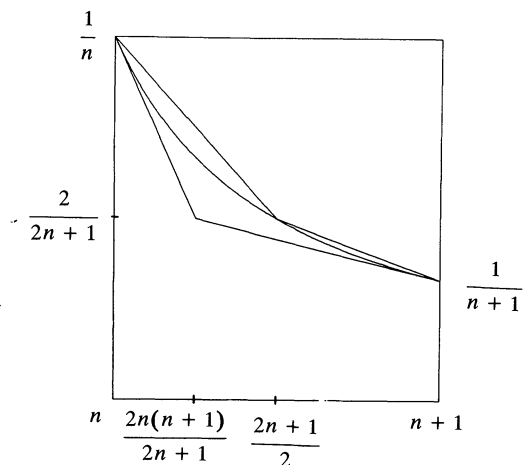


Figure 2

The weakest bounds are the basic elementary integral error bounds; the strongest bounds are newly computed by the author and are based on a suggestion in [1, p. 89].

Depending on the series in question, various bounds in (2) may give substantial, or only marginal, improvement over the use of other bounds also given in the above. For pedagogical purposes the author prefers the following inequality because of its clear improvement to the basic integral error bounds given in virtually all calculus texts. If $S = S_n + R_n$, papers [9] and [10] imply that

$$S_n + \int_{n+1}^{\infty} f + \frac{1}{2}f(n+1) < S < S_n + \int_n^{\infty} f - \frac{1}{2}f(n+1). \quad (3)$$

We adapt (3) to relationship (1) by letting \underline{S} , S and \bar{S} be the sums of three series for which $\underline{S} < S < \bar{S}$. Also let $\underline{f}(n)$, $\bar{f}(n)$, \underline{S}_n and \bar{S}_n be the generators and partial sums for \underline{S} and \bar{S} respectively. Then

$$\underline{S}_n + \int_{n+1}^{\infty} \underline{f} + \frac{1}{2}\underline{f}(n+1) < S < \bar{S}_n + \int_n^{\infty} \bar{f} - \frac{1}{2}\bar{f}(n+1). \quad (4)$$

Applying (4) to (1) and summing the telescoping series, we obtain

$$\begin{aligned} \frac{1}{4} + \frac{1}{2} \left\{ \sum_{k=1}^n \frac{1}{k(2k+1)} + \ln \frac{2n+3}{2n+2} + \frac{1}{(2n+2)(2n+3)} \right\} \\ < \gamma < \sum_{k=1}^n \frac{1}{k(2k+1)} + \ln \frac{2n+1}{2n} - \frac{1}{(2n+2)(2n+3)} \end{aligned} \quad (5)$$

This table easily follows.

n			
1	0.553238	0.688798	(all values are rounded)
2	0.555647	0.632667	
10	0.556824	0.614033	
49	0.556852	0.613709	
50	0.556853	0.613709	
77	0.556853	0.613707	
78	0.556853	0.613706	

After iteration 78 there is no further improvement; hence this method gives $0.55685 < \gamma < 0.61371$, with the lower bound conventionally truncated and the upper bound always rounded up.

Method 1 is less satisfactory than the almost 15,000 iterations necessary to generate four place rounded accuracy with Method 0. The following modification of the above technique, however, leads to bounds that converge to γ itself.

Convergence to γ , Method 1'. The difficulty with Method 1 is that the intervals (that generate those series giving lower and upper bounds for γ) are fixed at unit width, so that, on the interval $[n, n+1]$, the contributions to the bounds are poor when n is small. A way to remedy this is to replace the partial sums \underline{S}_n and \bar{S}_n in (4) by $\gamma_n = \sum_{k=1}^n 1/k - \ln(n+1)$. From FIGURE 2, we have $\underline{S}_n < \gamma_n < \bar{S}_n$, $\underline{R}_n < \gamma_R < \bar{R}_n$ (where $\gamma_R = \gamma - \gamma_n$), and, with (4) above, easily see that

$$\begin{aligned} \underline{S}_n + \int_{n+1}^{\infty} \underline{f} + \frac{1}{2}\underline{f}(n+1) &< \gamma_n + \int_{n+1}^{\infty} \underline{f} + \frac{1}{2}\underline{f}(n+1) < \gamma_n + \underline{R}_n < \gamma_n + \gamma_R \\ &= \gamma < \gamma_n + \bar{R}_n < \gamma_n + \int_n^{\infty} \bar{f} - \frac{1}{2}\bar{f}(n+1) \\ &< \bar{S}_n + \int_n^{\infty} \bar{f} - \frac{1}{2}\bar{f}(n+1) \end{aligned} \quad (6)$$

Consequently, the bounds given by that part of (6) which is

$$\gamma_n + \int_{n+1}^{\infty} f + \frac{1}{2}f(n+1) < \gamma < \gamma_n + \int_n^{\infty} \bar{f} - \frac{1}{2}\bar{f}(n+1) \quad (7)$$

will converge to γ . By combining (1) with (7) we obtain

$$\begin{aligned} & \frac{1}{4(n+1)} + \sum_{k=1}^n \frac{1}{k} - \ln(n+1) + \frac{1}{2} \ln \frac{2n+3}{2n+2} + \frac{1}{(4n+4)(2n+3)} \\ & < \gamma < \sum_{k=1}^n \frac{1}{k} - \ln(n+1) + \ln \frac{2n+1}{2n} - \frac{1}{(2n+2)(2n+3)}. \end{aligned}$$

From this we compute

n			γ	
1	0.568425	0.662318		
2	0.573701	0.600722		
9	0.576969	0.578069		
10	0.577014	0.577887	0.577	(truncated)
14	0.577111	0.577528		
15	0.577125	0.577483	0.577	(rounded)
35	0.577199	0.577257		
36	0.577200	0.577254	0.5772	(truncated)
107	0.577214	0.577220		
108	0.577214	0.577219	0.5772	(rounded)
355	0.577215	0.577216		
356	0.577216	0.577216	0.577216	(rounded)

In order to accelerate the convergence further, we could bisect the interval $[n, n+1]$ and repeat Method 1 for the subintervals $[n, n+1/2]$ and $[n+1/2, n+1]$, from which

$$\begin{aligned} & \frac{1}{8} \sum_{n=1}^{\infty} \left[\frac{1}{n} - \frac{1}{n+1} \right] + \frac{1}{4} \sum_{n=1}^{\infty} \frac{48n^2 + 44n + 9}{n(2n+1)(4n+1)(4n+3)} \\ & < \gamma < \sum_{n=1}^{\infty} \frac{8n+3}{n(4n+1)(4n+3)} \end{aligned}$$

which converge, however, to values strictly lower/higher than γ . We then employ Method 1', using the integral remainder bounds given by (4) and replacing the partial sums by γ , to get

$$\begin{aligned} & \frac{1}{8(n+1)} + \gamma_n + \frac{1}{4} \ln \frac{(2n+3)(4n+5)(4n+7)}{32(n+1)^3} \\ & + \frac{48n^2 + 140n + 101}{8(n+1)(2n+3)(4n+5)(4n+7)} \\ & < \gamma < \gamma_n + \frac{1}{2} \ln \frac{(4n+1)(4n+3)}{16n^2} - \frac{8n+11}{2(n+1)(4n+5)(4n+7)}. \end{aligned}$$

Convergence to a six-place (rounded) accurate value for γ is by the 183rd iterate. Informally, this appears to imply that the method is $O(h)$ in the step size (at least for this example).

Other Examples and Extensions. The techniques described in Methods 1 and 1' above can be used to evaluate the Euler constant for a *convergent* series and, therefore, to accelerate finding the value of the series itself. For example, consider $\sum_{n=1}^{\infty} 1/n^2$, whose value is $\pi^2/6$. The computations for the Euler constant, given by Method 1', converge to 0.644934 (rounded) by the 79th iteration. Adding a value of one thus gives an approximate value to the series. By contrast, calculations done in [5] and based on a slightly weaker form of (4) generate the same value by the 124th iteration. The above method thus converges faster, though at the expense of more analytical work.

Method 1 may also be used to accelerate the convergence of series that we do not normally associate with convergence tests for series of positive terms. For example, consider the alternating harmonic series $\sum_{n=1}^{\infty} (-1)^{n+1}/n = \ln 2$. We may rewrite this series in the form $\sum_{n=1}^{\infty} [1/(2n-1) - 1/2n] = \sum_{n=1}^{\infty} 1/[2n(2n-1)]$. [It is not by accident that we choose this example; its similarity to the series in (1) is clear. In fact, the bounds in (1) can be shown to be related to the alternating harmonic series and to have the exact values of $5/4 - \ln 2$ and $2 - 2\ln 2$, respectively.] When we apply (3) to the series in this form, we obtain

$$\sum_{k=1}^n \frac{1}{2k(2k-1)} + \frac{1}{2} \ln \frac{2n+2}{2n+1} + \frac{1}{(4n+4)(2n+1)}$$

$$< \ln 2 < \sum_{k=1}^n \frac{1}{2k(2k-1)} + \frac{1}{2} \ln \frac{2n}{2n-1} - \frac{1}{(4n+4)(2n+1)}.$$

Convergence to $\ln 2 = 0.693147$ (rounded) is by the 87th iterate. [If, instead, we use the strongest error bounds from (2) for a similar computation, we obtain the above value as early as the 25th iterate, a substantial improvement. In addition, see [7] for another approach to the evaluation of $\ln 2$ by using various other series to estimate remainders.]

Conclusions. It is to be hoped that, with the ease and availability of machine computing, a somewhat more sophisticated approach to the evaluation of series may take place in the current calculus classroom. Error bounds such as those given in (3) or, indeed, any mix or match of bounds given in (2) are within the grasp of the student. In addition, the *sec/tan method* (Method 1') for accelerating the convergence of the Euler constant perhaps merits some consideration because of its easily visualized geometry.

REFERENCES

1. R. P. Boas, Jr., Estimating Remainders, *Math Mag.*, 51 (1978) 83–89.
2. R. P. Boas, Jr., Partial sums of infinite series and how they grow, this MONTHLY, 84 (1977) 237–258.
3. R. P. Boas, Jr., Growth of partial sums of divergent series, *Math. of Comp.*, 31 (1977) 257–264.
4. R. P. Boas, Jr., and J. W. Wrench, Jr., Partial sums of the harmonic series, this MONTHLY, 78 (1971) 864–870.
5. R. J. Collins, Approximating series, *Coll. Math. J.*, 23 (1992) 153–157.
6. H. Flanders, *MicroCalc ver. 5.1*, MathCalcEduc, Ann Arbor, MI, 1990.
7. C. Goldsmith, Calculation of $\ln 2$ and π , *Math. Gaz.*, 55 (1971) 434–436.
8. G. H. Hardy, *Orders of Infinity*, 2nd ed., Cambridge Univ. Press, New York, 1924.
9. R. K. Morley, The remainder in computing by series, this MONTHLY, 57 (1950) 550–551; repr. in *Selected Papers on Calculus*, 1968, 324–325.
10. R. K. Morley, Further note on the remainder in computing by series, this MONTHLY, 58 (1951) 410–412.

11. D. A. Smith, *INTERFACE: Calculus and the Computer*, 2nd ed., Saunders, New York, 1984.
12. S. K. Stein and A. Barcellos, *Calculus and Analytic Geometry*, 5th ed., McGraw-Hill, New York, 1992.
13. G. B. Thomas and R. L. Finney, *Calculus and Analytic Geometry*, 7th ed., Addison-Wesley, Reading, MA, 1989.
14. S. R. Tims and J. A. Tyrrell, Approximate evaluation of Euler's Constant, *Math. Gaz.*, 55 (1971) 65–67.

*Penn State-Scranton
120 Ridge View Drive
Dunmore, PA 18512*

The Prisoner's Paradox Revisited

Awaiting the dawn sat three prisoners wary,
A trio of brigands named Tom, Dick and Mary.
Sunrise would signal the death knoll of two,
Just one would survive, the question was who.

Young Mary sat thinking and finally spoke.
To the jailer she said, "You may think this a joke"
But it seems that my odds of surviving 'til tea,
Are clearly enough just one out of three.

But one of my cohorts must certainly go,
Without question, that's something I already know.
Telling the name of one who is lost,
Can't possibly help me. What could it cost?"

The shriveled old jailer himself was no dummy,
He thought, "But why not?" and pointed to Tommy.
"Now it's just Dick and I" Mary chortled with glee,
"One in two are my chances, and not one in three!"

Imagine the jailer's chagrin, that old elf
She'd tricked him, or had she? Decide for yourself.

*Richard E. Bedient
Department of Mathematics
and Computer Science
Hamilton College
Clinton, NY 13323*

A Focusing Property of the Ellipse

Marc Frantz

INTRODUCTION. In this article some elementary mathematics will be used to establish and quantify a focusing property of ellipses and ellipsoids. This property is evidently unknown, even though it is quite simple and apparently significant. Possible areas of application include solar energy collection and laser technology; we also use our results to solve an open problem recently posed in this journal. It is a well-known fact from geometry that a light ray which leaves a focus F_1 of an ellipse will be reflected to the other focus F_2 . This is illustrated by the curve $F_1R_1F_2$ in FIGURE 1. The goal of this article is to describe the long-term behavior of such a light ray, and to describe the corresponding long-term behavior of an idealized spherical wave emitted from one focus of an ellipsoid of revolution.

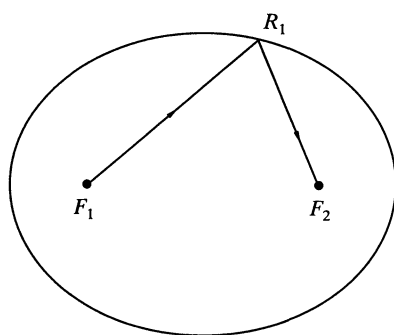


Figure 1

In a qualitative sense, the essential features of the “future history” of such a ray are depicted in FIGURE 2, where the ray $F_1R_1R_2R_3R_4R_5$ has been drawn until just before the fifth reflection at R_5 . Given the ellipse and foci, the ray is easy to

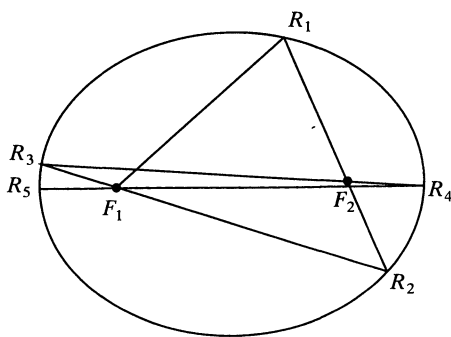


Figure 2

sketch by hand, since it must always travel from focus to focus. In the process of sketching such rays, one discovers that the same thing always happens to any ray, namely, its path eventually flattens out until it is essentially moving back and forth along the major axis of the ellipse.

To make this assertion more precise, we define the *departure angle from a focus*, of a ray which has just left that focus, to be the smallest positive angle subtended by the departing ray and a “reference ray” from that focus to the opposite focus. For example, the successive departure angles of the ray in FIGURE 2 are $\angle F_2 F_1 R_1$, $\angle F_1 F_2 R_2$, $\angle F_2 F_1 R_3$, $\angle F_1 F_2 R_4$, and $\angle F_2 F_1 R_5$, which is nearly equal to π . Thus departure angles range from 0 to π , and for each departure angle in $(0, \pi)$ there are two corresponding rays which can depart from a particular focus. This ambiguity will cause no problem for us, since what we wish to show is that the successive departure angles of almost any ray tend to π , just as they did in FIGURE 2.

Proposition 1. *Let a light ray leave a focus of an ellipse with departure angle $\theta_0 \in (0, \pi)$, and let the successive departure angles of the ray be $\theta_1, \theta_2, \dots$. Then $\theta_n \uparrow \pi$.*

Proof: From FIGURE 3 we see that if $\theta_n \in (0, \pi)$, then $\theta_n = \theta_{n+1} - \alpha_n < \theta_{n+1} < \pi$, and thus $\theta_{n+1} \in (0, \pi)$ and $\theta_n \uparrow$. Since the sequence $\{\theta_n\}$ is bounded, it must converge, and hence $\alpha_n \rightarrow 0$. By the law of sines applied to triangle $F_1 R_{n+1} F_2$, we then have

$$\lim_{n \rightarrow \infty} \frac{\sin \theta_n}{s_n} = \lim_{n \rightarrow \infty} \frac{\sin \alpha_n}{L} = 0.$$

Now the sequence of lengths s_n is clearly bounded, so $\sin \theta_n \rightarrow 0$, and since $\theta_n \uparrow$, this implies $\theta_n \uparrow \pi$.

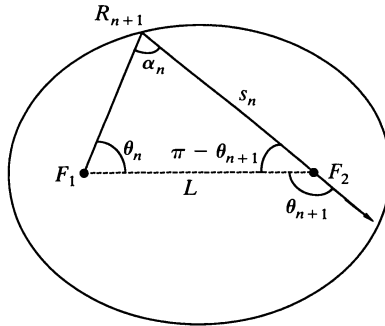


Figure 3

Remark. An immediate application of Proposition 1 is the solution of an open problem posed by J. E. Connett in [1]. The negative form of the question asked there is: Can a perfectly reflecting container be designed which will admit a beam of parallel light rays from some particular direction, but will not allow any ray from that beam to escape, regardless of how many times it is reflected? Such a counterexample is indeed possible, and a two-dimensional version of such an object appears in FIGURE 4.

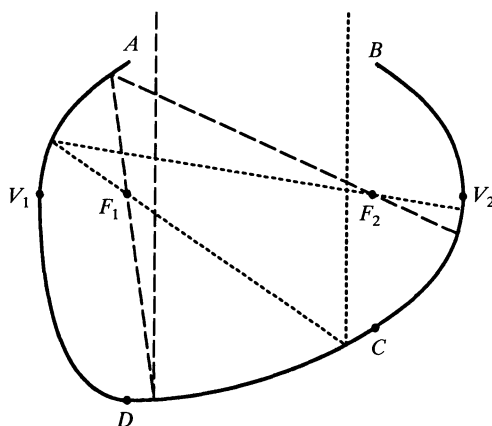


Figure 4. A light trap.

The smooth curve in the figure is constructed using a “primary ellipse” with equation $x^2/a^2 + y^2/b^2 = 1$, where $a > b > 0$, a parabola with equation $(x + c)^2 = 4a(y + a)$, where $c = \sqrt{a^2 - b^2}$, and a “secondary ellipse” with equation $(x + c)^2/(a - c)^2 + y^2/a^2 = 1$, which is used merely to make a smooth connection. The arcs V_1A and BV_2C belong to the primary ellipse, which has foci $F_1 = (-c, 0)$ and $F_2 = (c, 0)$. Arc CD belongs to the parabola, which has focus F_1 and vertex D , and arc DV_1 belongs to the secondary ellipse. The reader may check that the connecting points are: $V_1 = (-a, 0)$, $D = (-c, -a)$, $C = (c, -b^2/a)$, and that the arcs have common tangents at these points. Endpoints A and B belong to the curve and are directly above the focal points. Thus any light ray entering vertically must strike the parabola and be reflected to F_1 . The ray then strikes the upper left quadrant of the primary ellipse, and is reflected through F_2 to the lower right quadrant of the same ellipse. By Proposition 1 and the geometry of the figure, we see that subsequent reflection points are either in the upper left quadrant converging downward to vertex V_1 , or in the lower right quadrant converging upward to vertex V_2 . Hence the light ray gradually approaches a horizontal trajectory and never leaves the “container”. A three-dimensional version of the container is easily made by using the curve to generate a cylinder.

Let us now turn our attention to ellipsoids and examine the aspect of light intensity in terms of the convergence of $\{\theta_n\}$ described in Proposition 1. What we will see (Proposition 3) is that the intensity of a light wave emitted from a focus of an ellipsoid of revolution increases with each reflection at one particular point on the wave, called the point of concentration. This intensification, or focusing effect, is described by a simple, discrete-time exponential law which has apparently dramatic consequences. An illustration of this will be given at the end of the article.

To begin with, given any initial departure angle θ , it is possible to derive an explicit formula for the n th subsequent departure angle θ_n ; this formula is the tool that will enable us to give a precise description of the focusing process and to derive the exponential law mentioned above. To this end, consider the diagram in FIGURE 5, where we have drawn an arbitrary ellipse with eccentricity ε and major axis of length $2a$, with $d(F_1, F_2) = 2a\varepsilon$. A ray leaves F_1 with departure angle θ and travels a distance r before being reflected at R_1 . Using polar coordinates with

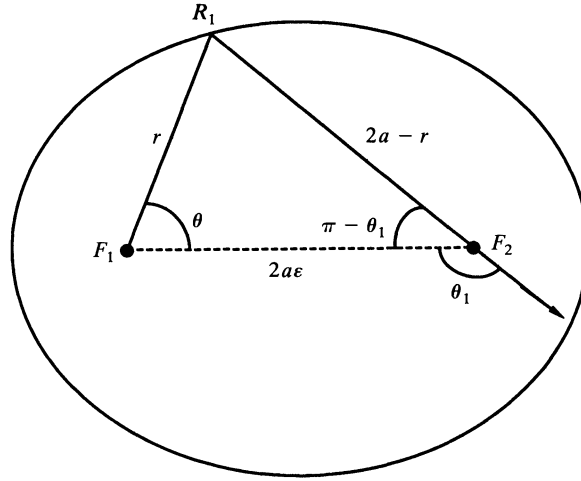


Figure 5

origin F_1 , we can determine that

$$(*) \quad r = \frac{a(1 - \varepsilon^2)}{1 - \varepsilon \cos \theta}.$$

It is well-known that the ray must travel a distance $2a - r$ to F_2 , where it arrives at an angle $\pi - \theta_1$ and departs at an angle θ_1 .

Let us now derive a formula for the departure angle θ_1 in terms of the departure angle θ . Applying the law of cosines to the triangle $F_1R_1F_2$, and using the fact that $\cos(\pi - \theta_1) = -\cos \theta_1$, we have

$$r^2 = (2a\varepsilon)^2 + (2a - r)^2 + 2(2a\varepsilon)(2a - r)\cos \theta_1.$$

The solution for $\cos \theta_1$ can be put in the form

$$\cos \theta_1 = \frac{1 - a(1 + \varepsilon^2)/r}{2a\varepsilon/r - \varepsilon}.$$

Substituting the expression in (*) for r gives

$$(**) \quad \cos \theta_1 = \frac{1 - \frac{(1 + \varepsilon^2)}{(1 - \varepsilon^2)}(1 - \varepsilon \cos \theta)}{\frac{2\varepsilon(1 - \varepsilon \cos \theta)}{(1 - \varepsilon^2)} - \varepsilon} = \frac{(1 + \varepsilon^2)\cos \theta - 2\varepsilon}{(1 + \varepsilon^2) - 2\varepsilon \cos \theta}.$$

To get a form of this equation which will be more useful to us, we now multiply the numerator and denominator of the last expression in (**) by $2/(1 - \varepsilon)^2$ to obtain

$$(***) \quad \cos \theta_1 = \frac{\frac{2(1 + \varepsilon^2)}{(1 - \varepsilon)^2} \cos \theta - \frac{4\varepsilon}{(1 - \varepsilon)^2}}{\frac{2(1 + \varepsilon^2)}{(1 - \varepsilon)^2} - \frac{4\varepsilon}{(1 - \varepsilon)^2} \cos \theta}.$$

Defining the constant μ by

$$\mu := \left(\frac{1 + \varepsilon}{1 - \varepsilon} \right)^2, \quad (1)$$

and noticing that $2(1 + \varepsilon^2)/(1 - \varepsilon)^2 = \mu + 1$ and $4\varepsilon/(1 - \varepsilon)^2 = \mu - 1$, we see that $(***)$ can be written

$$\cos \theta_1 = \frac{(\mu + 1)\cos \theta - (\mu - 1)}{(\mu + 1) - (\mu - 1)\cos \theta}.$$

Finally, defining the function $f_\mu: [0, \pi] \rightarrow [0, \pi]$ by $f_\mu(\theta) := \theta_1$, we can compute θ_1 by

$$\theta_1 = f_\mu(\theta) = \arccos \left(\frac{(\mu + 1)\cos \theta - (\mu - 1)}{(\mu + 1) - (\mu - 1)\cos \theta} \right) \quad (2)$$

since $\theta_1 \in [0, \pi]$.

If we use the standard notation $f_\mu^1(\theta) := f_\mu(\theta)$, $f_\mu^2(\theta) := f_\mu(f_\mu(\theta))$, \dots , then the successive departure angles of a ray are $\theta, f_\mu^1(\theta), f_\mu^2(\theta), \dots$. If we also adopt the conventions $f_\mu^0(\theta) := \theta$ and $f_\mu^{-n}(\theta) := (f_\mu^n)^{-1}(\theta)$, we have

Proposition 2. *Let $\mu > 1$, and let $f_\mu: [0, \pi] \rightarrow [0, \pi]$ be defined as in (2). Then for each $\theta \in [0, \pi]$ and each $n \in \mathbb{Z}$,*

$$f_\mu^n(\theta) = \arccos \left(\frac{(\mu^n + 1)\cos \theta - (\mu^n - 1)}{(\mu^n + 1) - (\mu^n - 1)\cos \theta} \right). \quad (3)$$

Proof: Since (3) obviously holds when $n = 0$ or $n = 1$, suppose that (3) is true when $n = k$ for some $k > 0$. By straightforward algebra it follows that for any $\theta \in [0, \pi]$,

$$f_\mu^{k+1}(\theta) = f_\mu^k(f_\mu(\theta)) = \arccos \left(\frac{(\mu^{k+1} + 1)\cos \theta - (\mu^{k+1} - 1)}{(\mu^{k+1} + 1) - (\mu^{k+1} - 1)\cos \theta} \right),$$

and hence (3) is true for any $n \geq 0$. It is equally straightforward to check that for any $n \geq 0$, defining f_μ^{-n} according to (3) gives

$$f_\mu^{-n}(f_\mu^n(\theta)) = f_\mu^n(f_\mu^{-n}(\theta)) = \theta$$

for each $\theta \in [0, \pi]$. Thus each f_μ^n is invertible, and (3) is proved.

Remarks. It is not difficult to use (3) to prove Proposition 1. For example, if $\theta \in (0, \pi)$, then by elementary calculus, $\lim_{n \rightarrow \infty} f_\mu^n(\theta) = \pi$. It is also clear from (3) that each f_μ^n is a homeomorphism of $[0, \pi]$ with fixed points 0 and π ; we shall use this fact later. Both of these results are suggested by the graphs in FIGURE 6.

Before giving a precise statement of the promised exponential law, it will help to look at some pictures and see how the focusing process evolves. In FIGURE 7 are 21 cells from a movie, and these should be viewed from left to right, starting at the top. In the first cell (marked A), 16 particles simultaneously depart from the left-hand focus F_1 , all at the same speed. The particles have a uniform angular spacing, and thus the initial departure angles are $0, \pi/8, 2\pi/8, \dots, \pi$, where the angles $\pi/8, \dots, 7\pi/8$ each correspond to a *pair* of particles which belong to opposite half-planes determined by the major axis.

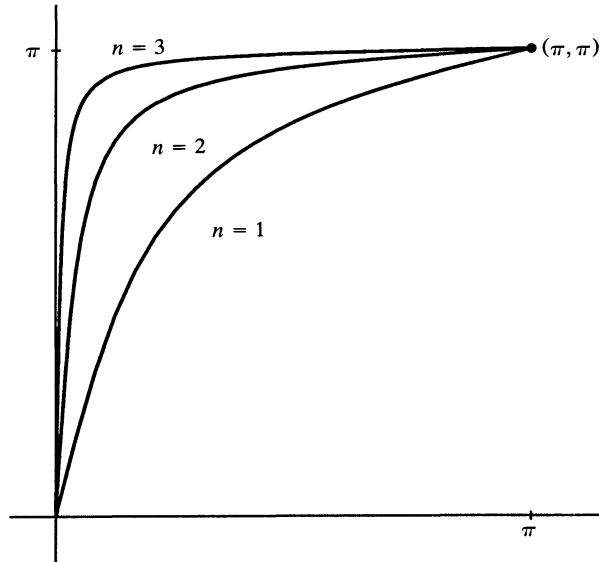


Figure 6. Graphs of $f_{l_0}^n(\theta)$ for $n = 1, 2, 3$.

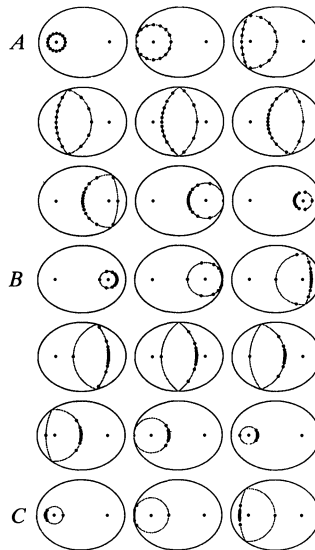


Figure 7

Initially, the particles “ride” on a circular wave, but as the wave begins to be reflected, it consists of two circular arcs: one centered at, and diverging from, F_1 , and the other centered at, and converging toward, the other focus F_2 . This follows from the fact that the total path length from focus to focus is the same for each particle, and thus the particles must eventually converge on F_2 in a circular pattern. The particles on the diverging arc become more dispersed, while those on the converging arc become more concentrated.

By the end of the third row of cells, the wave has become a small circle converging on F_2 , and in cell B the wave has crossed itself (we allow the “particles” to pass through each other) and is now diverging from F_2 . The departure angles, now being first iterates of f_μ , are mostly clustered about π (recall that at F_2 , the departure angle π is to the *right*). As the wave diverges from F_2 , the particles at first begin to disperse, but very soon most of them are reflected and begin to converge back toward F_1 .

In cell C , the particles have just passed through F_1 , and their departure angles are computed by $f_\mu^2(\theta)$. Except for the particle which had initial departure angle $\theta = 0$, the particles are now contained within an arc of about 70° . Let us observe what this process implies for a spherical wave:

A spherical light wave of uniform intensity, emitted from one focus of a perfectly reflecting, prolate ellipsoid of revolution, will eventually be reduced to a single, highly concentrated pulse oscillating back and forth along the major axis.

Of course, when the wave is emitted, there should be no emitting device left at the focus, or else it would interfere with the focusing process. Thus one might imagine two thin wires with a tiny gap at F_1 , where a spark starts the process. Perhaps better still, one might imagine the wave to be emitted from a matter-anti-matter collision at F_1 .

Having visualized this process, we are now ready for the precise statement of the exponential law. Let us assume that we have a prolate ellipsoid of revolution with $\mu > 1$, and that a sphere of unit radius centered at either focus can be contained inside the ellipsoid. For the purpose of analysis we will consider only spherical waves of unit radius departing alternately from F_1 and F_2 , as roughly depicted in cells A , B , and C of FIGURE 7. On any such wave, let us refer to the point farthest from the opposite focus (departure angle π) as the *point of concentration*. By the *intensity* at a point of the wave we mean the power (energy per unit time) per unit area transported by the wave at that point. What we will show is that, if the initial wave has uniform intensity I_0 , then the intensity I_n at the point of concentration after the n th reflection is given by

$$I_n = I_0 \mu^n. \quad (4)$$

Thus μ is an excellent measure of the focusing efficiency of an ellipsoid. Moreover, from (1) we see that $\mu \uparrow \infty$ as $\varepsilon \uparrow 1$, so elongated ellipsoids focus more efficiently.

Proposition 3. *Let $\mu > 1$ and let I_0 and I_n be as defined above. Then for each $n \geq 0$, I_n is given by (4).*

Proof: Let us think of all the departing spherical waves of unit radius described above as being superimposed on one another, with their points of concentration at the left. Then a small spherical cap (a special case of a *zone*, shaded in profile in FIGURE 8) centered at this point, with its boundary located at θ , has area $A_0 = 2\pi(1 + \cos \theta)$. This follows from a standard theorem of solid geometry which says that a zone with altitude h on a sphere of radius r has area $2\pi rh$; in this case, $r = 1$ and $h = 1 + \cos \theta$.

Since for any $n \geq 1$, the function f_μ^n is a homeomorphism of $[0, \pi]$, we may also consider it to be a homeomorphism of the sphere which increases the θ -coordinates of points, where $\theta \in (0, \pi)$ is a polar angle measured from the extreme right-hand point of the sphere. Thus after n reflections, a larger spherical cap centered at π will be mapped by f_μ^n onto the smaller one. The boundary of this

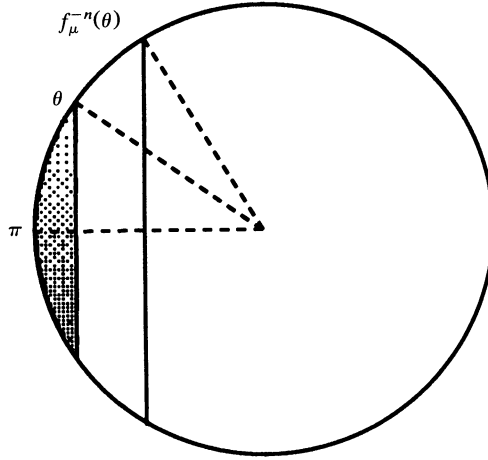


Figure 8

larger cap is located at $f_\mu^{-n}(\theta)$, and hence by the zone formula and (3) its area is

$$A_{-n} = 2\pi \left[1 + \cos(f_\mu^{-n}(\theta)) \right] = 2\pi \left[1 + \frac{(\mu^{-n} + 1)\cos \theta - (\mu^{-n} - 1)}{(\mu^{-n} + 1) - (\mu^{-n} - 1)\cos \theta} \right].$$

Now the original intensity I_0 was constant, so the energy flux through the larger cap was initially $I_0 A_{-n}$. After this cap is mapped by f_μ^n onto the smaller one, the total energy flux is the same (the same rays now pass through the smaller cap), but the area has decreased to A_0 . Hence the average intensity over the cap is now $I_0 A_{-n}/A_0$. The intensity I_n at the point of concentration is then given by

$$\begin{aligned} I_n &= \lim_{\theta \rightarrow \pi} \frac{I_0 A_{-n}}{A_0} \\ &= I_0 \lim_{\theta \rightarrow \pi} \frac{(\mu^{-n} + 1) - (\mu^{-n} - 1)\cos \theta + (\mu^{-n} + 1)\cos \theta - (\mu^{-n} - 1)}{[(\mu^{-n} + 1) - (\mu^{-n} - 1)\cos \theta](1 + \cos \theta)} \\ &= I_0 \lim_{\theta \rightarrow \pi} \frac{2 + 2\cos \theta}{(\mu^{-n} + 1) + 2\cos \theta - (\mu^{-n} - 1)\cos^2 \theta}. \end{aligned}$$

L'Hospital's rule is clearly applicable, and thus

$$\begin{aligned} I_n &= I_0 \lim_{\theta \rightarrow \pi} \frac{-2\sin \theta}{-2\sin \theta + 2(\mu^{-n} - 1)\cos \theta \sin \theta} \\ &= I_0 \lim_{\theta \rightarrow \pi} \frac{-2}{-2 + 2(\mu^{-n} - 1)\cos \theta} = I_0 \mu^n. \end{aligned}$$

To summarize our results, we may consider both the traditional focusing process and the one considered here as means of increasing the *intensity* of a wave emitted from a focus of an ellipsoid of revolution. The former process makes the intensity arbitrarily large over the entire wave, in a *periodic* fashion, by periodically making the wave arbitrarily small (compressing it to a focus). The latter process makes the intensity arbitrarily large at one particular point on the wave, in an *exponential* fashion, by reflecting the wave arbitrarily many times. Of course, this process is exponential in the discrete-time sense, since we have fixed our attention on a

particular phase of the wavefront cycle (i.e. departing from either focus with unit radius), which occurs every $2a/v$ seconds, assuming that the major axis has length $2a$ and light travels at v units per second. In between reflections and focal-point compressions, the intensity increases or decreases according to an inverse-square law, depending on whether that part of the wave is converging to or diverging from a focus. However, we may consider the wavefront at *any* time t , except when the wave is compressed to a point, identify the point of concentration, and show that at time $t + 2a/v$ the intensity at the corresponding point has increased by a factor of μ . Thus it seems appropriate to call this process “Phase-Exponential Reflection,” and to call our make-believe antimatter-powered device a “Phase-Exponential Reflector” (PHASER). Under the idealized conditions we have described, PHASER intensification of light is rather dramatic, even for moderately elongated ellipsoids. To illustrate this, we will look at one final example.

In FIGURE 9 we have sketched an ellipsoid of revolution with eccentricity $\varepsilon = 24/25$. There also appears in that figure a wave of radius 1 departing from F_1 . For this ellipsoid, $\mu = 2401$. Hence assuming that the wave initially departs from F_1 with uniform intensity I_0 , we can use (4) to determine that when the wave next departs from F_1 (after two reflections), the intensity at the point of concentration will be increased by a factor of almost six million.

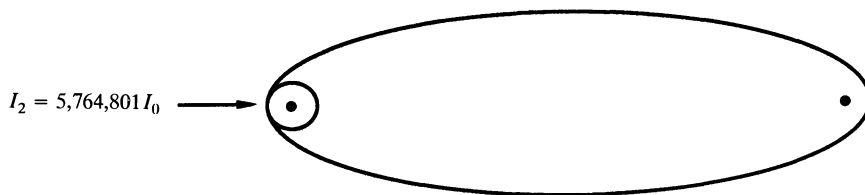


Figure 9

The author is indebted to Professor Y. Ionin for an improvement of the original version of the proof of Proposition 1; to Professor Richard Patterson for a helpful critique; and to the referees for several suggestions and improvements.

REFERENCE

1. J. E. Connett, Trapped reflections?, *Amer. Math. Monthly*, 99 (1992) 178–179.

Department of Mathematical Sciences
IUPUI
1125 East 38th Street
Indianapolis, IN 46205-2820

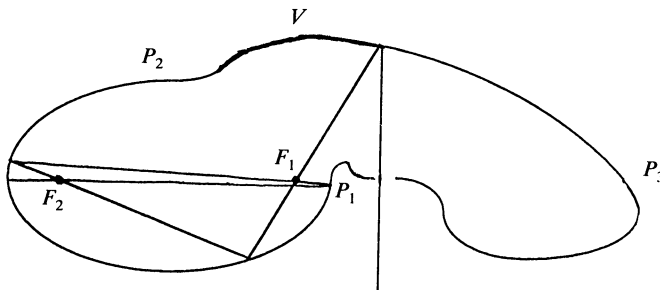
NOTES

Edited by: John Duncan

Reflections Can Be Trapped

Roberto Peirone

In [1] it is asked whether, roughly speaking, a beam of light rays can be trapped in an opportune container, acting as a reflecting mirror. In this paper I prove that this is so. More precisely, I describe a Jordan curve of class C^1 (in fact, we can make this curve to be of class C^∞), $\gamma:[0, 1] \rightarrow \mathbb{R}^2$, such that $\gamma([\frac{1}{2}, 1]) = A$ is a horizontal open segment line, and, by supposing that $\gamma([0, \frac{1}{2}))$ works as a reflecting mirror, every vertical upward light ray passing through A enters the bounded region determined by $\gamma([0, 1])$ and will not repass through A . Analogously, and this example is more closely related to the problem treated in [1], we can construct a compact connected surface ς of class C^1 in \mathbb{R}^3 , such that there exists an open rectangle A contained in ς , and, by supposing that $\varsigma \setminus A$ works as a reflecting mirror, every light ray passing through A , orthogonal to A and entering the bounded region determined by ς , will not repass through A . The precise statement of the problem discussed in [1] is whether in this type of situation, in fact in a slightly more general situation, it is possible to prove that some of the light rays of the beam will be reflected back out through A . As pointed out by the referee, it may be shown that if we allow just a minor deformation from a parallel ray, then it is impossible to trap the beam. The idea of the construction is to join opportunely an ellipse to a parabola whose focus is one of the foci of the ellipse (see FIGURE).



The curve in the FIGURE represents $\gamma([0, \frac{1}{2}))$. The arc between P_1 and P_2 is an arc of the ellipse whose foci are F_1 and F_2 . The arc between V and P_3 is an arc of the parabola whose vertex is V and whose focus is F_1 . Here, the straight-line passing through F_1 and F_2 is horizontal and that passing through V and F_1 is vertical. If $\gamma([0, \frac{1}{2}))$ works as a reflecting mirror, then every vertical upward light ray as in the FIGURE will remain in the interior of the curve. For, after the first reflection, by a well-known property of the parabola, it will pass through F_1 , after

the second reflection, by a well-known property of the ellipse, it will pass through F_2 , then it will re-pass through F_1 , then through F_2 and so on.

In order to construct s as required it is sufficient, for example, to consider a surface opportunely built on a cylinder having γ for base.

REFERENCE

1. J. E. Connett, Trapped Reflections?, *Amer. Math. Monthly* 99 (1992) 178–179.

*Dipartimento di Matematica
II Università di Roma
Via della Ricerca Scientifica
00133 Roma (Italy)
Peirone@mat.utovrm.it*

Euler's Theorem

Katherine Heinrich and Peter Horak

Fermat's Little Theorem. *If p is a prime and $p \nmid a$, then $a^{p-1} \equiv 1 \pmod{p}$.*

Stated in a letter by Fermat in 1640, the first public proof of this result was given by Euler in 1736 (although it appears that Leibniz also had a proof which he did not publish). The details of the situation are described in [Dic]. Some time later (in 1760) Euler generalized Fermat's result.

Euler's Theorem. *If a and m are relatively prime, then $a^{\phi(m)} \equiv 1 \pmod{m}$, where $\phi(m) = |\{n: 1 \leq n \leq m \text{ and } n \text{ and } m \text{ are relatively prime}\}|$.*

Since then a variety of proofs have been presented. The most frequently cited depending on the use of either the binominal theorem, group theory, or number theoretic arguments [see Dic]. We present a proof which is combinatorial in spirit. (When this proof is restricted to the case when m is prime, Thue's proof [Thu] of Fermat's Little Theorem is obtained.)

Proof of Euler's Theorem. We first prove the theorem in the case when m is a prime power; $m = p^\alpha$, p prime. Let $S_a^m = \{(a_1, a_2, \dots, a_m): 1 \leq a_i \leq a\}$ be a set of a^m ordered m -tuples. Considering the mapping σ acting on S_a^m as follows. For $A = (a_1, a_2, \dots, a_m)$, $\sigma(A) = (a_m, a_1, a_2, \dots, a_{m-1})$. That is, σ cycles the entries of A one place to the right. Now we are going to count the number of orbits of S_a^m , under the action of σ , of the maximum size m . This will be done by counting the total number of elements which are members of smaller orbits. Suppose $A \in S_a^m$ belongs to an orbit consisting of $t < m$ elements. Then $\sigma^t(A) = A$ and $t|m$ (or $t = p^\beta$, $\beta \leq \alpha - 1$). We can therefore express A as

$$A = (a_1, a_2, \dots, a_t, a_1, a_2, \dots, a_t, \dots, a_1, a_2, \dots, a_t),$$

and hence as

$$A = (a_1, a_2, \dots, a_r, a_1, a_2, \dots, a_r, \dots, a_1, a_2, \dots, a_r),$$

where $r = p^{\alpha-1}$. On the other hand, whenever A has the form

$$A = (a_1, a_2, \dots, a_r, a_1, a_2, \dots, a_r, \dots, a_1, a_2, \dots, a_r), \quad \sigma^r(A) = A.$$

Therefore, there are $(a^m - a^r)/m$ orbits of size m . Thus $m|a^m - a^r$ and as $(m, a) = 1$, $m|a^{m-r} - 1$. But $a^{m-r} - 1 = a^{p^\alpha - p^{\alpha-1}} - 1 = a^{\phi(m)} - 1$ and therefore $a^{\phi(m)} \equiv 1 \pmod{m}$.

Now let $m = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_t^{\alpha_t}$, p_i prime, $1 \leq i \leq t$. Since $\phi(m) = (p_1^{\alpha_1} - p_1^{\alpha_1-1})(p_2^{\alpha_2} - p_2^{\alpha_2-1}) \cdots (p_t^{\alpha_t} - p_t^{\alpha_t-1})$ and $p_i^{\alpha_i} | a^{\phi(m)} - 1$, then $m | a^{\phi(m)} - 1$ or $a^{\phi(m)} \equiv 1 \pmod{m}$.

ACKNOWLEDGMENT. The research was done while the second author was visiting Simon Fraser University; he would like to thank the Department of Mathematics and Statistics for its hospitality.

REFERENCES

- [Dic] L. E. Dickson, History of the Theory of Numbers, Chelsea Pub. Com, New York, 1919.
 [Thu] A. Thue, Ein kombinatorischer Beweis eines Satzes von Fermat, Videnskabs-Selskabers Skrifter, Math.-Naturv. Klasse (1910) 1–7.

*Dept. of Mathematics and Statistics
 Simon Fraser University
 Burnaby, British Columbia
 Canada V5A 1S6*

*Southern Illinois University
 Carbondale, IL 62901*

PICTURE PUZZLE (from the collection of Paul Halmos)



To some people he is known as his brother's brother.
 (see page 265.)

Fix a (d, n) -labeled graph G and vertices u and v . Consider a logspace Turing machine that uses the contents of the random tape to perform a walk of length $4n^4d$ on G starting at u . If the Turing machine ever reaches v it returns 0, otherwise it returns 1. When the contents of the random tape is truly random the Turing machine returns 1 with probability at most $1/4n^2$. If the random tape is actually the result of some pseudorandom generator the probability should still be close to $1/4n^2$, else this logspace Turing machine could distinguish the pseudorandom inputs from the random ones. Thus the Turing machine with pseudorandom input will return 1 with probability at most, say, $1/2n^2$. Since there are at most n^2 choices for u and v , the probability that the pseudorandom generator produces a sequence that does not cover G is at most $1/2$. This means that every (d, n) -labeled graph is covered by at least half of the possible outputs of the generator, so the concatenation of all of the outputs must be (d, n) -universal. Since there are $2^{O(\log^2 n)}$ possible inputs, and each output has polynomial length, the concatenation of all outputs has length $n^{O(\log n)}$.

Further Reading. Wigderson [9] provides an extensive overview of the s - t connectivity problem, universal traversal sequences, L vs. NL, and many other related combinatorial and algorithmic problems.

REFERENCES

1. R. Aleliunas, R. Karp, L. Lovász, R. Lipton, and C. Rackoff, *Random Walks, Universal Traversal Sequences, and the Complexity of Maze Problems*, Proceedings of the 20th Symposium on Foundations of Computer Science, IEEE Computer Society, Los Alamitos, 1979, pp. 218–223.
2. L. Babai, N. Nisan, and M. Szegedy, *Multiparty Protocols, Pseudorandom Generators for Logspace, and Time-Space Trade-Offs*, Journal of Computer and System Sciences, 45 (1992), pp. 204–232.
3. M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, San Francisco, 1979.
4. S. Itrail, *Polynomial Universal Traversal Sequences for Cycles are Constructible*, Proceedings of the 20th Symposium on Theory of Computing, ACM, New York, pp. 491–503.
5. H. Karloff, R. Paturi, J. Simon, *Universal Traversal Sequences of Length $n^{O(\log n)}$ for Cliques*, Information Processing Letters, 28 (1988), pp. 241–243.
6. B. Marion, “The Computer Science Sampler: Turing Machines and Computational Complexity,” *The American Mathematical Monthly*, January 1994.
7. N. Nisan, *Pseudorandom Generators for Space-Bounded Computation*, Combinatorica, 12 (1992), pp. 449–461.
8. W. Savitch, *Relationships Between Nondeterministic and Deterministic Tape Complexities*, Journal of Computer and System Sciences, 4 (1970), pp. 177–192.
9. A. Wigderson, *The Complexity of Graph Connectivity*, Proceedings of the 17th Mathematical Foundations of Computer Science Conference, Lecture Notes in Computer Science, vol. 629, eds.: Havel and Koubek, Springer, Berlin, 1992, pp. 112–132.

AT & T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974
jf@research.att.com
reingold@research.att.com

Answer to Picture Puzzle (p. 261)

Harald Bohr, the brother of physicist Niels Bohr.

Universal Traversal Sequences

Joan Feigenbaum and Nick Reingold

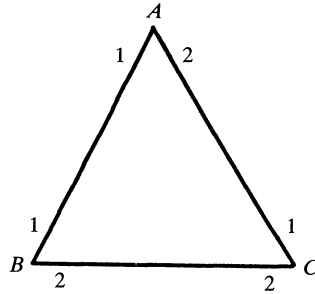
In this article we discuss a purely combinatorial problem, the construction of short universal traversal sequences, and its relationship to questions about logspace computation. We state the problem formally, show how it arises naturally in complexity theory, and review some of the known partial results. A basic introduction to complexity theory can be found in [6].

The P vs. NP problem is recognized by the mathematical world as the central open question in the theory of computation. Less widely known outside of computer science is the fact that the analogous question for space-bounded computation was resolved long ago¹: Savitch [8] shows that any language accepted by a nondeterministic Turing machine that uses space $O(s(n))$ is also accepted by a deterministic Turing machine that uses space $O((s(n))^2)$. Hence deterministic polynomial space is equivalent (in terms of language-recognition power) to nondeterministic polynomial space: PSPACE = NPSPACE.

However, one question about the relationship of nondeterministic space-bounded computation and its deterministic counterpart remains open: Is the quadratic “blow-up” in space complexity exhibited by Savitch’s construction necessary? This question turns out to be most interesting for computations that use very little space. Let L be the class of languages accepted by deterministic Turing machines using only $O(\log n)$ space, and NL be the class of languages accepted by nondeterministic Turing machines using only $O(\log n)$ space. It could be the case that Savitch’s theorem is optimal at this low end of the space-complexity spectrum. On the other hand, it could be that any language recognizable by a nondeterministic Turing machine is also recognizable by a deterministic Turing machine with the same space complexity; if that’s true, then NL is exactly equal to L . The truth might also lie somewhere between these two extremes.

Universal Traversal Sequences. We now consider a purely combinatorial problem. Consider d -regular, undirected graphs $G = (V, E)$. Such a graph is called (d, n) -labeled if it has n vertices and the edges incident to each vertex are labeled by a permutation of $\{1, 2, \dots, d\}$. Note that an edge has two labels, one associated with each of its endpoints, and that these labels may differ. A sequence of labels $\sigma = \sigma_1 \sigma_2 \dots \sigma_m$ and a *start vertex* s define a *walk* $s_0 s_1 \dots s_m$ as follows. Let $s_0 = s$. For $1 \leq i \leq m$, if $\sigma_i = j$, then s_i is the (unique) vertex such that there is an edge $e = \{s_{i-1}, s_i\}$ and the label of e that is associated with s_{i-1} is j . For example, in the

¹by computer science standards.



(2, 3)-labeled graph above, the sequence 12211 and start vertex A define the walk ABCBAB.

A sequence σ is called (d, n) -universal if, for every connected (d, n) -labeled graph (V, E) and every start vertex $s \in V$, the walk defined by σ and s contains every vertex in V . For example, the sequence 112 is $(2, 3)$ -universal.

For expository purposes, we also define a directed version of universal traversal sequences. A d -regular directed graph $G = (V, A)$ is called (d, n) -labeled if it has n vertices and the arcs out of each vertex are labeled with a permutation of $\{1, 2, \dots, d\}$. Note that, in the directed case, arcs have only one label. A sequence σ and start vertex s define a walk in essentially the same way as they do in the standard definition. We say that σ is (d, n) -directed-universal if, for every strongly connected (d, n) -labeled digraph (V, A) and every start vertex s , the walk defined by σ and s contains every vertex in V .

The connection between universal traversal sequences and L vs. NL is made via the s - t connectivity problem. An instance of this problem consists of a directed graph G and two vertices s and t in $V(G)$. The question is whether there is a path in G from s to t . There is a straightforward nondeterministic logspace algorithm for this problem: Guess a path $s_0 s_1 s_2 \dots s_{n-1}$ such that $s = s_0$ and then check that each arc $s_{i-1} \rightarrow s_i$ is present in G and that some $s_i = t$. This algorithm is clearly correct, and it only requires space $O(\log n)$, where $n = |V(G)|$. To see why so little space is required, note that the algorithm does *not* need to store the entire path at any time. Rather, it need only store the names of two consecutive vertices s_{i-1} and s_i ; once it has verified the presence of the arc $s_{i-1} \rightarrow s_i$, it can write over s_{i-1} with its guess for s_{i+1} .

The language STCONN of yes-instances of the s - t connectivity problem is in fact NL-complete. Furthermore, STCONN remains NL-complete if we assume that the input digraphs are regular. So, if we could exhibit a deterministic logspace algorithm for (regular) s - t connectivity, we would have shown that L is equal to NL. It would suffice to exhibit, for each constant d , a Turing machine that uses $O(\log n)$ space and generates a sequence τ_n that is (d, n) -directed-universal. Note that the restriction to $O(\log n)$ space implies that the length of τ_n is polynomial in n . Unfortunately, it can be shown that no family $\{\tau_n\}_{n \geq 1}$ of polynomial-length directed-universal sequences exists. (L might still be equal to NL, but the equality cannot be proven this way.) This raises the question of the existence of a family $\{\sigma_n\}_{n \geq 1}$ that is deterministically logspace-generable such that σ_n is (d, n) -universal (which is weaker than (d, n) -directed universal). Aleliunas *et al.* [1] give a beautiful probabilistic argument that, for any d , there is a polynomial-length family $\{\sigma_n\}_{n \geq 1}$ such that σ_n is (d, n) -universal. Whether such a family can be generated in deterministic logspace remains open.

Constructing Universal Traversal Sequences. Consider a random walk in a connected undirected graph. At each time step the next vertex is chosen uniformly from the neighbors of the current vertex. We say the walk *covers* G if every vertex is visited at least once during the walk. For any vertex v , let C_v be the expected time at which a random walk starting at v covers G . The maximum over all v of C_v is called the *cover time* of the graph. Aleliunas *et al.* [1] show that the cover time for any d -regular graph with n vertices is at most n^2d . They then use this observation to prove the existence of (d, n) -universal traversal sequences of polynomial length as follows.

Let σ be a sequence of labels of length $4n^3d^2 \log_2 nd$ chosen uniformly from the set of all such sequences. We will show that the probability that σ is (d, n) -universal is not zero. The probability that σ is not (d, n) -universal is the same as the probability that there exists a (d, n) -labeled graph G and a vertex v such that a random walk of length $4n^3d^2 \log_2 nd$ starting at v does not cover G . This is the same as the probability that there exists a (d, n) -labeled graph G and a vertex v such that G is not covered by $2nd \log_2 nd$ consecutive random walks, each of length $2n^2d$, the first of which is started at v . Since any graph G has cover time at most n^2d , Markov's inequality shows that a random walk of length $2n^2d$, starting from any vertex, has probability at most $1/2$ of *not* covering G . If we take $2nd \log_2 nd$ such random walks consecutively, the probability that none of them covers G is at most $(1/2)^{2nd \log_2 nd} = (nd)^{-2nd}$. Thus, for any fixed G and v , the walk through G starting at v given by σ has probability at most $(nd)^{-2nd}$ of not covering G . There are at most $(nd)^{nd}$ choices for v and G , so summing over all these choices shows that the probability that σ is not universal is strictly less than one.

This proves the existence of polynomial-length universal traversal sequences and suggests the possibility that the language USTCONN (the yes-instances of the s - t connectivity problem for undirected graphs) is in L . However, the above proof does not show this, since it does not show how to *generate* the traversal sequences using only $O(\log n)$ space. Whether this is possible is still an open problem. There are two interesting partial results that are worth mentioning. For $d = 2$, Istrail [4] gives a construction of polynomial-length traversal sequences, but his sequences cannot be constructed in deterministic logspace. For $d = n - 1$, Karloff *et al.* [5] give an explicit construction of traversal sequences of length $n^{O(\log n)}$.

Traversal Sequences and Pseudorandom Generators. The best explicit universal traversal sequences constructed so far are due to Nisan [7]. This construction exploits a connection, due to Babai *et al.* [2], between traversal sequences and *pseudorandom generators*. In this context, a pseudorandom generator is an algorithm that converts a small number of truly random bits into a long sequence of bits that appears random to any Turing machine that uses only a limited amount of space. We will not give a precise definition of “appears random to any Turing machine that uses only a limited amount of space.” The interested reader should see [2] for details. Nisan's generators convert a truly random string of length $O(S \log R)$ into a string of length R that appears random to any Turing machine that uses space at most S . In particular, if $S = O(\log n)$ and R is polynomial in n , then Nisan's generators can convert $O(\log^2 n)$ truly random bits into polynomially many bits that appear random to any Turing machine that uses only $O(\log n)$ space. We will now show that the concatenation of all the possible outputs of Nisan's pseudorandom generator is a (d, n) -universal traversal sequence of length $n^{O(\log n)}$.

Fix a (d, n) -labeled graph G and vertices u and v . Consider a logspace Turing machine that uses the contents of the random tape to perform a walk of length $4n^4d$ on G starting at u . If the Turing machine ever reaches v it returns 0, otherwise it returns 1. When the contents of the random tape is truly random the Turing machine returns 1 with probability at most $1/4n^2$. If the random tape is actually the result of some pseudorandom generator the probability should still be close to $1/4n^2$, else this logspace Turing machine could distinguish the pseudorandom inputs from the random ones. Thus the Turing machine with pseudorandom input will return 1 with probability at most, say, $1/2n^2$. Since there are at most n^2 choices for u and v , the probability that the pseudorandom generator produces a sequence that does not cover G is at most $1/2$. This means that every (d, n) -labeled graph is covered by at least half of the possible outputs of the generator, so the concatenation of all of the outputs must be (d, n) -universal. Since there are $2^{O(\log^2 n)}$ possible inputs, and each output has polynomial length, the concatenation of all outputs has length $n^{O(\log n)}$.

Further Reading. Wigderson [9] provides an extensive overview of the s - t connectivity problem, universal traversal sequences, L vs. NL, and many other related combinatorial and algorithmic problems.

REFERENCES

1. R. Aleliunas, R. Karp, L. Lovász, R. Lipton, and C. Rackoff, *Random Walks, Universal Traversal Sequences, and the Complexity of Maze Problems*, Proceedings of the 20th Symposium on Foundations of Computer Science, IEEE Computer Society, Los Alamitos, 1979, pp. 218–223.
2. L. Babai, N. Nisan, and M. Szegedy, *Multiparty Protocols, Pseudorandom Generators for Logspace, and Time-Space Trade-Offs*, Journal of Computer and System Sciences, 45 (1992), pp. 204–232.
3. M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, San Francisco, 1979.
4. S. Itrail, *Polynomial Universal Traversal Sequences for Cycles are Constructible*, Proceedings of the 20th Symposium on Theory of Computing, ACM, New York, pp. 491–503.
5. H. Karloff, R. Paturi, J. Simon, *Universal Traversal Sequences of Length $n^{O(\log n)}$ for Cliques*, Information Processing Letters, 28 (1988), pp. 241–243.
6. B. Marion, “The Computer Science Sampler: Turing Machines and Computational Complexity,” *The American Mathematical Monthly*, January 1994.
7. N. Nisan, *Pseudorandom Generators for Space-Bounded Computation*, Combinatorica, 12 (1992), pp. 449–461.
8. W. Savitch, *Relationships Between Nondeterministic and Deterministic Tape Complexities*, Journal of Computer and System Sciences, 4 (1970), pp. 177–192.
9. A. Wigderson, *The Complexity of Graph Connectivity*, Proceedings of the 17th Mathematical Foundations of Computer Science Conference, Lecture Notes in Computer Science, vol. 629, eds.: Havel and Koubek, Springer, Berlin, 1992, pp. 112–132.

AT & T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974
jf@research.att.com
reingold@research.att.com

Answer to Picture Puzzle (p. 261)

Harald Bohr, the brother of physicist Niels Bohr.

THE EVOLUTION OF ...

Edited by Abe Shenitzer

Mathematics, York University, North York, Ontario M3J 1P3, Canada

What Are Algebraic Integers and What Are They For?

John Stillwell

My title is stolen from Dedekind's *Was sind und was sollen die Zahlen?* (1888), the classic book on the meaning of the natural numbers. His book is actually about set theory rather than number theory or algebra but, as we shall see, Dedekind had a lot more up his sleeve. His reflections on the algebraic properties of the natural numbers would have filled a much larger book, which unfortunately he never had time to write. The following is a sketch of the book that might have been.

The natural numbers $0, 1, 2, 3, \dots$ are the setting for the oldest and deepest mathematical problems. As early as 300 B.C., Euclid recognised that *divisibility* and *primes* are important concepts for natural numbers, and that even simple propositions about them involve subtle methods of proof. For example, to prove the seemingly obvious proposition that if a prime p divides the product of natural numbers a, b then p divides a or p divides b , Euclid had to introduce the concept of *greatest common divisor* (gcd), and use the euclidean algorithm to express the gcd in a suitable form. It follows almost immediately from this proposition about prime divisors that each natural number has a *unique prime factorisation*. That is, if n is a natural number, then n can be expressed as a product of primes in exactly one way, up to the order of factors. Unique prime factorisation was isolated as the “fundamental theorem of arithmetic” in the classic *Disquisitiones Arithmeticae* of Gauss (1801).

Gauss organised much of earlier number theory into a cohesive structure by introducing unifying ideas, such as congruence, and proving unifying theorems, such as quadratic reciprocity. But Gauss also burst the bounds of the structure he had created with an array of new and baffling theorems. For example, he proved several theorems about the n th roots of 1, bearing on the geometric construction of regular n -gons. He showed in passing that these *irrational* numbers had application to the natural numbers; they were even connected with quadratic reciprocity (*Disquisitiones*, section 356). How come? Gauss's successors spent most of the 19th century searching for an answer to this question, for concepts to explain the effectiveness of irrational numbers in arithmetic. The most important of these was the concept of algebraic integer.

What algebraic integers are for. Actually, Gauss was not the first to use irrational numbers to answer questions about the natural numbers. Euler's *Algebra* (1770) contains several examples, the simplest being a proof of the following claim of Fermat: *among natural numbers, 27 is the only cube that exceeds a square by 2*. This

claim is one of the famous marginal notes in Fermat's copy of Bachet's *Diophantus*, made around 1637. Fermat claimed to have a rigorous proof, but never revealed it. Euler's proof is not exactly rigorous, but it is distinguished by the wild idea of using the number $\sqrt{-2}$.

The problem amounts to showing that $x = 5, y = 3$ is the only solution of

$$y^3 = x^2 + 2.$$

Euler factorises the right hand side into $(x + \sqrt{-2})(x - \sqrt{-2})$, then confidently assumes that $x + \sqrt{-2}, x - \sqrt{-2}$ must both be cubes, since their product is a cube and he can show that their gcd is 1. This implies, presumably, that

$$x + \sqrt{-2} = (a + b\sqrt{-2})^3$$

for some integers a, b , and hence

$$x + \sqrt{-2} = a^3 - 6ab^2 + (3ab^2 - 2b^3)\sqrt{-2}.$$

Equating imaginary parts we get

$$1 = 3a^2b - 2b^3 = b(3a^2 - 2b^2)$$

which implies $b = \pm 1$ and $3a^2 - 2b^2 = \pm 1$ since $+1$ and -1 are the only integer divisors of 1. It follows that $a = \pm 1$, which gives $x = 5$ as the only positive integer value of $a^3 - 6ab^2$. Q.E.D!

Well OK, we don't know what these things $a + b\sqrt{-2}$ and their divisors really are, or how they work, but they seem amazingly good for answering questions about the natural numbers. The same goes for other irrational objects arising from the factorisation of polynomials. Euler gave many examples, the most spectacular being a proof of Fermat's last theorem for cubes:

$$x^3 + y^3 = z^3 \text{ has no solution in positive integers}$$

(Fermat himself had settled the case of fourth powers by his method of infinite descent). He proved this using the factorisation

$$x^3 = z^3 - y^3 = (z - y) \left(z - \frac{-1 + \sqrt{-3}}{2} y \right) \left(z - \frac{-1 - \sqrt{-3}}{2} y \right)$$

and arguing, as above, that each factor in the right hand side must be a cube. This time the irrational objects involved are numbers of the form $a + b\omega$, where

$$\omega = \frac{-1 + \sqrt{-3}}{2}$$

is a complex cube root of 1. Gauss's work on the n th roots of 1 could be seen as a generalisation of this, involving objects of the form $a_0 + a_1\zeta + \cdots + a_{n-2}\zeta^{n-2}$, where ζ is a complex n th root of 1 and a_0, a_1, \dots, a_{n-2} are ordinary integers.

Thus by 1800 it was clear that irrational objects were good for number theory. What was not clear was the nature of these objects and the reasons they behaved like integers, if indeed they did. A lot more work had to be done before the right concept of "algebraic integer" was isolated.

Algebraic integers. Since we want to do arithmetic on algebraic integers, they must satisfy the requirement:

1. *The algebraic integers are closed under $+$, $-$ and \times .*

Also, since the aim in extending the concept of integer is to answer questions about ordinary integers, the relations between ordinary integers had better not

change in the process of extension. In particular, an ordinary integer a should not become a divisor of an ordinary integer b if it is not one already. That is, the rational number b/a should not be an algebraic integer unless it is an ordinary integer. This gives us a second requirement:

2. *The rational algebraic integers are the ordinary integers.*

The numbers $a + b\sqrt{-2}$ satisfy these requirements, as do the numbers $a_0 + a_1\zeta + \cdots + a_{n-2}\zeta^{n-2}$. The only problematic operation is multiplication of numbers $a_0 + a_1\zeta + \cdots + a_{n-2}\zeta^{n-2}$, as this creates powers $\zeta^{n-1}, \zeta^n, \dots$. The trick is to rewrite $\zeta^{n-1}, \zeta^n, \dots$ as combinations of $1, \zeta, \dots, \zeta^{n-2}$ with ordinary integer coefficients, using the equation

$$1 + \zeta + \cdots + \zeta^{n-2} + \zeta^{n-1} = 0$$

satisfied by ζ . Then a product of numbers $a_0 + a_1\zeta + \cdots + a_{n-2}\zeta^{n-2}$ is seen to be another number of the same form.

Now the numbers $\sqrt{-2}$ and ζ arise in the first place as the solutions of polynomial equations with ordinary integer coefficients, namely

$$x^2 + 2 = 0, \quad x^{n-1} + \cdots + x + 1 = 0.$$

Is the same true for the combinations $a + b\sqrt{-2}$ and $a_0 + a_1\zeta + \cdots + a_{n-2}\zeta^{n-2}$? The answer is yes, and even more is true: the combinations satisfy *monic* polynomial equations with ordinary integer coefficients. This follows from a general property of monic polynomial equations pointed out by Eisenstein (1850).

If $f(x) = 0$ is a monic polynomial equation with ordinary integer coefficients and roots $\alpha_1, \alpha_2, \dots, \alpha_n$, and if $g(\alpha_1, \dots, \alpha_n)$ is any polynomial in the roots with ordinary integer coefficients, then $g(\alpha_1, \dots, \alpha_n)$ is also the root of a monic polynomial equation $h(x) = 0$ with ordinary integer coefficients.

His proof applies Newton's theorem on symmetric polynomials to the product $h(x)$ of all terms $x - g(\alpha_{\sigma(1)}, \dots, \alpha_{\sigma(n)})$, where σ is a permutation of $1, \dots, n$, to conclude that the coefficients of $h(x)$ are ordinary integers. Obviously $h(x)$ is monic, and its roots include $g(\alpha_1, \dots, \alpha_n)$. Since $\alpha_1 + \alpha_2, \alpha_1 - \alpha_2$ and $\alpha_1\alpha_2$ are particular cases of $g(\alpha_1, \dots, \alpha_n)$, closure under $+$, $-$ and \times will hold when algebraic integers are defined as follows.

Definition. An *algebraic integer* is the root of a monic polynomial equation with ordinary integer coefficients.

Just as easily, this definition meets the second requirement:

A rational algebraic integer is an ordinary integer.

Because if a/b is a rational number (in lowest terms) that satisfies the equation

$$x^n + c_1x^{n-1} + \cdots + c_{n-1}x + c_n = 0$$

with ordinary integer coefficients, then substitution and rearrangement give

$$a^n/b = -c_1a^{n-1} - \cdots - c_{n-1}ab^{n-2} - c_nb^{n-1},$$

which is possible only if $b = 1$, since the right hand side is an ordinary integer.

Eisenstein did not actually state the definition or the closure under $+$, $-$ and \times , but he probably didn't need to. The class of solutions of monic polynomial equations was under consideration by other mathematicians at the time, and its

basic properties were implicit in known results about polynomials. In particular, the nature of rational algebraic integers was well known as a result about polynomials 50 years earlier (Gauss assumed it without proof in article 11 of the *Disquisitiones*). More important was Eisenstein's recognition that the algebraic integers shared properties with the ordinary integers. This pointed the way to a systematic explanation of the phenomena discovered by Euler and Gauss.

Divisors and primes. Closure of the algebraic integers under $+$, $-$ and \times means that they inherit the ring properties of the complex numbers. Hence they are a ring, like the ordinary integers, and satisfy the same basic propositions such as the commutative, associative and distributive laws. Unfortunately, many useful propositions are not logical consequences of these laws and in fact are *false* in the larger ring. For example, there are no primes in the ring of all algebraic integers, because every algebraic integer α has the factorisation $\alpha = \sqrt{\alpha} \sqrt{\alpha}$, and $\sqrt{\alpha}$ is also an algebraic integer. This difficulty can be avoided by working in smaller rings tailored to particular problems, such as the ring of integers $a + b\sqrt{-2}$ where Euler looked for solutions of $y^3 = x^2 + 2$. "Sufficiently small" rings of algebraic integers do have primes, and also prime factorisation, but the big question is whether the factorisation is unique.

The first ring of algebraic integers to be examined in this light was

$$\mathbf{Z}[i] = \{a + bi : a, b \in \mathbf{Z}\},$$

the ring of *Gaussian integers*. Gauss (1832) showed that $\mathbf{Z}[i]$ has a divisibility theory like that of \mathbf{Z} , including unique prime factorisation, thanks to an analogue of the following *division property* of \mathbf{Z} . If $a, b \in \mathbf{Z}$ and $b \neq 0$ then there are $q, r \in \mathbf{Z}$ ("quotient" and "remainder") with $a = qb + r$ and $0 \leq |r| < |b|$. The analogous division property of $\mathbf{Z}[i]$ is that for any $\alpha, \beta \in \mathbf{Z}[i]$ with $\beta \neq 0$ there are $\mu, \rho \in \mathbf{Z}[i]$ with $\alpha = \mu\beta + \rho$ where $0 \leq |\rho| < |\beta|$ and the absolute value $|\cdot|$ is now distance in the complex plane.

The division property of $\mathbf{Z}[i]$ can be seen by viewing the Gaussian integer multiples of β as the corners of a lattice of squares in the plane. A typical square is the one with corners at $0, \beta, i\beta, (1+i)\beta$. The "remainder" ρ is simply the difference between α and the nearest corner in the lattice, and $|\rho| < |\beta|$ because the distance from any point in a square to the nearest corner is less than the length of a side. With the division property established, one has a euclidean algorithm for gcd, and the rest of the route to unique prime factorisation is the same as in \mathbf{Z} .

A similar geometric argument establishes a division property for $\mathbf{Z}[\sqrt{-2}]$. Hence $\mathbf{Z}[\sqrt{-2}]$ also has a unique prime factorisation, and Euler's treatment of $y^3 = x^2 + 2$ is valid. Euler's proof that $x^3 + y^3 \neq z^3$ can likewise be justified by finding a euclidean algorithm for $\mathbf{Z}[\omega]$.

Alas, it is not always as simple as this. Unique prime factorisation *fails* in $\mathbf{Z}[\sqrt{-5}]$, and it also fails in $\mathbf{Z}[\zeta]$ for n th roots ζ of unity from $n = 23$ onwards (Kummer, 1844). The failure in $\mathbf{Z}[\sqrt{-5}]$ was encountered implicitly by Fermat in studying primes of the form $x^2 + 5y^2$ (see [2], p. 82). The failure in $\mathbf{Z}[\zeta]$ was a major stumbling block, though not the only one, in Kummer's attempt to prove Fermat's last theorem.

Algebraic numbers and functions. The story of how unique factorisation was lost, and then regained by Kummer and Dedekind's theory of ideals, has often been told (for example in [1], pp. 818–824). I shall not repeat it here. Instead I shall sketch how Dedekind built the theory of algebraic integers to make unique factorisation possible, and what he discovered in the process.

Dedekind's main idea was to embed rings of algebraic integers in algebraic number *fields*, where concepts of linear algebra come to the surface. "Sufficiently small" rings of algebraic integers lie in fields K of finite dimension over \mathbf{Q} , and any such field is of the form $K = \mathbf{Q}(\alpha)$ where the degree of α equals the dimension of K . The concept of field and the existence of the "primitive element" α were already implicit in Abel and Galois, but Dedekind made the crucial identification of degree with dimension by observing that $\{1, \alpha, \alpha^2, \dots, \alpha^{n-1}\}$ is a basis for K , where n is the degree of α . This enabled him to give simple linear algebra proofs of theorems previously dependent on properties of symmetric functions, such as the closure of algebraic integers under $+$, $-$ and \times . It also enabled him to develop a "linear" approach to Galois theory, and to apply Galois theory to number theory.

These ideas were presented in the supplements Dedekind wrote for the 2nd, 3rd and 4th editions of Dirichlet's *Zahlentheorie* (1871, 1878, 1893). They were extended in Hilbert's *Zahlbericht* (1897) and presented there in almost modern form. Many 20th century algebraists, such as Emmy Noether, Artin and van der Waerden, learned their algebra from these works. In fact, Emmy Noether used to say "Es steht schon bei Dedekind" ("It's already in Dedekind").

A great benefit to flow from Dedekind's approach to algebraic numbers was a new approach to algebraic *functions*. Among algebraic functions of one variable, the polynomials play the role of the integers. This idea goes back at least as far as Stevin's *L'arithmétique* (1585), where the euclidean algorithm is used to find the gcd of polynomials. Dedekind took the idea that polynomials are "integers", and generalised to an analogy of algebraic integers. The ring $\mathbf{C}[z]$ of complex polynomials in z is extended to the field $\mathbf{C}(z)$ of rational functions, and each algebraic function lies in a finite-dimensional extension K of $\mathbf{C}(z)$. The functions analogous to algebraic integers are the *entire* functions—those with a finite value for each value of z . By carrying over concepts from number theory to function theory, Dedekind and Weber (1882) were able to give a completely algebraic definition of a *Riemann surface*. In doing so, they put many theorems about Riemann surfaces (that is, algebraic curves) on a sound basis for the first time, and laid the foundation of modern algebraic geometry.

This is very interesting, but is it number theory? Can this rarified form of geometry tell us anything about the ordinary integers? A full answer would cover most of 20th century mathematics, but the short answer of course is yes. All of these ideas, and much more, are needed for Andrew Wiles' proof of Fermat's last theorem.

REFERENCES

1. M. R. Kline, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, New York.
2. A. Weil, *Number Theory: An approach through History*, Birkhäuser, Boston.

Department of Mathematics
Monash University
Clayton 3168
AUSTRALIA

THE AUTHORS

FERNANDO GOUVÊA was born in Brazil and received his undergraduate and master's degrees there, at the University of São Paulo. He then went to Harvard University, where he got his doctorate in 1987 with a thesis on p -adic modular forms written under the direction of Barry Mazur. Since then, he has taught at the University of São Paulo, at Queen's University in Kingston, Ontario, and at Colby College in Waterville, ME, where he is assistant professor. Fernando's research interests are in number theory. Most of his work deals with p -adic modular forms and their associated Galois representations, but he has also been involved with elliptic curves and with diagonal hypersurfaces over finite fields. Besides research, Fernando also enjoys teaching mathematics, and relishes the challenge of writing understandable expository accounts of sophisticated and difficult work. He is a member of the AMS and MAA, and is part of a support network for minority students interested in Mathematics, Science, and Engineering organized by the New England Board of Higher Education. He has also been known to lurk in various *usenet* groups, to sing in the choir and lead Bible studies in his local church, and to write reviews of science fiction books for various small publications. Fernando lives in Waterville, Maine with his wife, two sons, and a dog.

DAVID ALDOUS received a Ph.D. from Cambridge University in 1977. His current research interests incline toward "modern discrete probability," indirectly motivated by computer science study of randomized algorithms and probabilistic analysis of deterministic algorithms. Since 1979 he has taught at U.C. Berkeley in the Department of Statistics, where he likes to annoy his colleagues by saying he is interested in the applications of probability to all scientific fields *except statistics*.

MICHAEL A. B. DEAKIN studied at the Universities of Melbourne and Chicago. He specialises in both Biomathematics and the History of Mathematics and has taught in Australia, the US, Papua New Guinea and Indonesia. He also edits a journal of School Mathematics and the present article grew from a request by one of that journal's readers for information on Hypatia.

RICHARD BARSHINGER earned his Ph.D. (part time) in mathematical sciences at SUNY-Binghamton (1981), while continuing to hold fulltime academic employment at Penn State-Scranton. [He would not recommend this method of approach to anyone.] His thesis advisor was Jim Geer, and he continues to work in the field of uniform asymptotics. In addition, he is a professional organist/harpsichordist, having studied antique Flemish harpsichords in Antwerp, while on a grant from the government of Belgium.

MARC FRANTZ received his B.F.A. in 1975 from the Herron School of Art in Indianapolis. After a subsequent 13-year career in painting and picture framing, which included 8 years of self-taught courses in mathematics and physics, he was admitted to the Purdue Graduate Program in Mathematics at IUPUI, where he received his M.S. in 1990. A paper he wrote during his first year of graduate school, *On Sierpiński's nonmeasurable set*, was published in Volume 139 of *Fundamenta Mathematicae*. He is currently Lecturer in Mathematics at IUPUI. His research interests, though varied, lean towards ideas which can be clearly and meaningfully visualized. The main result of this article was discovered while sketching one evening.

JOAN FEIGENBAUM received a B.A. in Mathematics from Harvard and a Ph.D. in Computer Science from Stanford. She is currently a member of the Computing Principles Research Department of AT&T Bell Laboratories in Murray Hill, NJ. Her research interests include computational complexity theory, cryptography and security, and graph theory and applications.

NICK REINGOLD received a B.A. in Mathematics from the University of Chicago and a Ph.D. in Computer Science from Yale. He is currently a member of the Computing Principles Research Department of AT&T Bell Laboratories in Murray Hill, NJ. His research interests include structural complexity theory and on-line algorithms.

JOHN STILLWELL studied at Melbourne University and MIT, and started teaching at Monash University in 1970. Since 1980 he has made a career of writing books on subjects he failed to understand as a student—topology, noneuclidean geometry and, most recently, algebra. Almost invariably, it seems, the necessary understanding can be found by reading the masters and reconstructing the history of the subject in modern language. Number theory is the next topic he plans to attack in this manner.

MICHAEL S. MAHONEY teaches history of science and technology at Princeton University, where he earned his Ph.D. in 1967. The author of a variety of studies on mathematics from Antiquity through the seventeenth century, including *The Mathematical Career of Pierre Fermat* (Princeton, 1973), he has more recently turned to the history of computing, with articles on “The History of Computing in the History of Technology” (*Annals of the History of Computing*, 1988) and “The Roots of Software Engineering” (*CWI Quarterly*, 1990), and is currently completing a book on the formation of theoretical computer science as a mathematical discipline, a portion of which he delivered as a paper to a Joint AMS/MAA Session on History of Mathematics in January 1992.



Let G_n be the undirected graph whose vertices are the unlabeled graphs on n vertices (e.g., G_4 has 11 vertices), two of which are adjacent in G_n if and only if one can be obtained from the other by deleting an edge.

(a) Show that neither G_4 nor G_5 contain Hamiltonian paths.

(b)* Does G_n contain a Hamiltonian path for any $n > 5$?

10371. *Proposed by Emil Yankov Stoyanov, Antiem I Mathematical School, Vidin, Bulgaria.*

Let B' and C' be points on the sides AB and AC , respectively, of a given triangle ABC , and let P be a point on the segment $B'C'$. Determine the maximum value of

$$\frac{\min\{[BPB'], [CPC']\}}{[ABC]}$$

where $[F]$ denotes the area of F .

10372. *Proposed by Paul R. Chernoff and Jacob Feldman, University of California, Berkeley, CA.*

Let $\langle f_n \rangle_1^\infty$ be a sequence of non-negative integrable functions on the unit interval $[0, 1]$. Write $\int_0^1 f_n(x) dx = c_n$ and suppose that $\sum c_n < \infty$.

(a) Suppose also that $\sum \sqrt{c_n} < \infty$. Show that there is a convergent series of non-negative terms a_n such that, for almost all $x \in [0, 1]$, $f_n(x) \leq a_n$ for all sufficiently large n .

(b) Show that the conclusion of (a) may fail if $\sum \sqrt{c_n} = \infty$.

10373. *Proposed by M. J. Pelling, Balliol College, Oxford, England.*

Let $E_n \subseteq I = [0, 1]$ be a sequence of measurable sets in the unit interval with measures $mE_n \geq \delta > 0$ bounded away from zero. Prove that there is a subsequence $\langle E_{n_i} \rangle$ whose intersection has the cardinality of the continuum.

10374. *Proposed by David L. Book, University of Maryland, College Park, MD.*

Given an integer N , characterize the smallest square in the plane containing N lattice points.

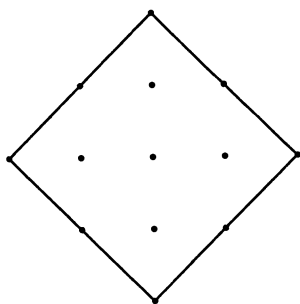
NOTES

Notes: (10373) Paul Halmos “Large Intersections of Large Sets”, this MONTHLY 99 (1992), 307–312, has a discussion of several related questions. (10374) The proposer supplied an extended narrative which is reproduced below. On that basis, it is proposed to call such a configuration a *Bolyai square*. Consideration of particular values of N should illustrate all possibilities of the position of a Bolyai square on the lattice.

Narrative on 10374. In the introduction to his translation of J. Bolyai's "The Science of Absolute Space" (*Non-Euclidean Geometry*, Ed. Roberto Bonola, Dover, New York, 1955), George Bruce Halsted wrote

"The profound mathematical ability of Bolyai János showed itself physically not only in his handling of the violin, where he was a master, but also of arms, where he was unapproachable... Bolyai, when in garrison with cavalry officers, was provoked by thirteen of them and accepted all of their challenges on condition that he be permitted after each duel to play a bit on his violin. He came out the victor from his thirteen duels, leaving his thirteen adversaries on the square."

Although Halsted neglected to mention it, Bolyai caused his victims to fall in the elegant and symmetric pattern shown below:



After concluding this bravura feat (as a result of which, not surprisingly, he was cashiered and lost his captain's commission in the Austrian army), Bolyai was heard to remark that the square in question is the smallest one containing 13 points of a square lattice.

SOLUTIONS

Another Appearance of Stirling Numbers

10188 [1992, 60]. *Proposed by Jürgen Eckhoff, University of California, Davis, CA.*

Define a function h on the non-negative integers by putting $h(0) = 0$ and $h(r) = \max\{i: i(i-1)/2 < r\}$ for $r \geq 1$. Let $f(r) = r + h(r) + 1$. For $1 \leq k \leq n$, let $T(n, k)$ be the number of integer sequences $0 = r_0 < r_1 < \cdots < r_{n-k}$ such that $f(r_{i-1}) \leq r_i \leq (k+i)(k+i-1)/2$ for $i = 1, 2, \dots, n-k$. Prove that $T(n, k)$ equals the Stirling number of the second kind $S(n, k)$.

Solution I by Ivar Skau, Telemark College, Bø, Norway. We prove that $T(n, k)$ and $S(n, k)$ satisfy the same recurrence. The Stirling number $S(n, k)$ is the

number of partitions of an n -set into k non-empty parts and satisfies the recurrence $S(n, k) = \sum_{r=k-1}^{n-1} \binom{n-1}{r} S(r, k-1)$, where the partitions are grouped by how many elements are in blocks different from the n th element. The initial condition here is $S(n, 1) = 1$ for $n \geq 1$. Note also that $S(n, k) = 0$ if $n < k$.

Let $\langle r_i \rangle$ be one of the sequences counted by $T(n, k)$; call these (n, k) -sequences. Let $H(s) = \left\{ r: \binom{s}{2} < r \leq \binom{s+1}{2} \right\}$; note that $|H(s)| = s$. If $r_{i-1} \in H(s)$, then $h(r_{i-1}) = s$, and hence $r_i \notin H(s)$. Indeed, if r_{i-1} is the j th element of $H(s)$, then r_i cannot be any of the first j elements of $H(s+1)$. These observations imply that for $k = 1$, the only legal sequence is the sequence with $r = \binom{i+1}{2}$. We also have $T(n, k) = 0$ if $n < k$. Hence the initial conditions for S and T agree.

For $k > 1$, we obtain a recurrence for $T(n, k)$. Let $R_k(n, i)$ be the number of (n, k) -sequences with $r_{n-k} = \binom{n-1}{2} + i$, for $1 \leq i \leq n-1$. The comments above about (n, k) -sequences and the sets $H(s)$ have several implications:

- 1) $R_k(n, n-1) = T(n-1, k)$, because an (n, k) -sequence ending with $\binom{n}{2}$ requires $r_{n-k-1} \leq \binom{n-1}{2}$.
- 2) $R_k(n, 1) = T(n-2, k-1)$, because an (n, k) -sequence ending with $\binom{n-1}{2} + 1$ requires $r_{n-k-1} \leq \binom{n-2}{2}$.
- 3) $R_k(n, i) - R_k(n, i-1) = R_k(n-1, i-1)$, because an $(n-1, k)$ -sequence extends to an (n, k) -sequence by appending $\binom{n-1}{2} + i$ but not $\binom{n-1}{2} + i-1$ if and only if $r_{n-k-1} = \binom{n-2}{2} + i-1$.

If we rewrite (3) as $R_k(n, i) = R_k(n, i-1) + R_k(n-1, i-1)$, we can iterate the recurrence $i-1$ times to obtain

$$R_k(n, i) = \sum_{j=0}^{i-1} \binom{i-1}{j} R_k(n-j, 1) = \sum_{j=0}^{i-1} \binom{i-1}{j} T(n-j-2, k-1),$$

where the last equality employs (2). If we set $i = n-1$, apply (1), and set $r = n-2-j$, this becomes

$$T(n-1, k) = R_k(n, n-1) = \sum_{r=0}^{n-2} \binom{n-2}{r} T(r, k-1).$$

Since $T(r, k-1) = 0$ when $r < k-1$, we can truncate the summation to begin at $r = k-1$. Now, since the choice of n was arbitrary, we find that $T(n, k)$ satisfies the same recurrence and initial conditions as $S(n, k)$, so $T(n, k) = S(n, k)$.

Solution II by Gruia Călinescu (student), University of Bucharest, Bucharest, Romania. We establish a bijection g from the set S of partitions of an n -set into k blocks to the set T of (n, k) -sequences. We turn a k -partition P of $[n] = \{1, \dots, n\}$ into an (n, k) -sequence $0 < r_1 < r_2 < \dots < r_{n-k}$, where the elements r_i are positions in an ordering on pairs of distinct positive integers. Let L be the reverse lexicographic ordering on these pairs, with $(2, 1) < (3, 1) < (3, 2) < (4, 1) < \dots$. To describe the positions in L , we write $1 = L(2, 1)$, $2 = L(3, 1)$, etc. In the notation of Solution I, note that $L(b, a) \in H(b-1)$, and hence $h(L(b, a)) = b-1$.

Beginning with the partition Q of $[n]$ into singletons, we iteratively combine two blocks that are contained within the same block of P . After doing this $n-k$ times, we will have P . We also maintain an order on the blocks of the current

partition Q ; in the initial Q , the j th block is $\{j\}$. With $i = n - k$ initially, let i denote the number of steps remaining, so Q has $k + i$ blocks.

Let C be the last block of Q that is not yet a block of P , and let B be the last block of Q that appears together with C in a single block of P . Let (s, t) be the positions of C, B in Q , and let $r_i = L(s, t)$. To form the next Q , replace C by $C \cup B$ in Q and eliminate B . The sets that were after B in Q (including $C \cup B$ in place of C) have their position in Q reduced by 1.

Since Q has $k + i$ blocks, $r_i \leq \binom{k+i}{2}$. Also $r_1 > 0$. We show $r_i \geq f(r_{i-1})$ to prove that $\{r_i\}_{i=1}^{n-k}$ is an (n, k) -sequence. Suppose $r_i = L(b, a)$ and $r_{i-1} = L(d, c)$. Then $f(r_{i-1}) = L(d, c) + d$. When we compute r_i , every block after the b th block of Q is already a block of P , so we have $d < b$. Similarly, if $d = b - 1$, then $c < a$. If $d < b - 1$, then $L(b, a) \geq L(d, c) + d = f(r_{i-1})$, since the length of $H(d + 1)$ is d . If $d = b - 1$, then $L(b, a) = L(d, c) + d - 1 + a - c \geq f(r_{i-1})$.

Hence our map g produces an (n, k) -sequence. The map is injective; it is easy to compute from distinct partitions a point at which the resulting sequences will differ. To show g is surjective, we construct the inverse. Starting with Q being the ordered partition into singletons, we determine for $i = n - k, \dots, k + 1$ the unique blocks that could be combined to produce the next partition while producing r_i . Since $r_i \leq \binom{k+i}{2}$, the pair (b, a) with $r_i = L(b, a)$ is a 2-element set with $a < b \leq k + i$. Hence we combine the a th block and b th block of Q . After $n - k$ steps, we obtain a partition P with k blocks.

The fact that $r_i \geq f(r_{i-1})$ for each i allows us to verify that these unions are precisely the unions performed when g is applied to P . It implies that the index of the higher combined block always decreases. Also, if that index decreases by exactly one, then the index of the lower combined block decreases. Hence a block remains unchanged once the two combined blocks have lower indices than it, and therefore it remains when P is reached. These facts imply that the choices are the same as when g is used to compute the sequence from P .

Solved also by J. H. Steelman and the proposer.

An Identity for the Other Stirling Numbers

10195 [1992, 162]. *Proposed by Andrew Granville, University of Georgia, Athens, GA.*

For $m \geq k \geq 1$, define numbers $b(k, m)$ by

$$b(1, m) = 1 \quad \text{for } m \geq 1,$$

$$b(k + 1, m) = \sum_{j=k}^{m-1} b(k, j) \left(\frac{1}{j} + \frac{1}{m-j} \right) \quad \text{for } m \geq k + 1 \geq 2.$$

For example, $b(2, m)$ is twice the $(m - 1)$ th partial sum of the harmonic series (for $m \geq 2$).

(a) Prove that

$$\sum_{k=1}^m \frac{(-1)^k}{k!} b(k, m) = 0 \quad \text{for } m \geq 2.$$

(b) Prove that $(m - 1)!b(k, m) = k!\sigma(m, k)$, where $\sigma(m, k)$ is the unsigned Stirling number of the first kind.

Composite solution by Robin J. Chapman, University of Exeter, Exeter, U.K., and Anchorage Math Solutions Group, University of Alaska, Anchorage, AK. Recall that $\sigma(m, k)$ is the number of permutations of $[m]$ having k cycles. We prove first that (a) follows from (b). Expressing $b(k, m)$ in terms of $\sigma(m, k)$ converts the desired identity to $(1/(m-1)!)\sum_{k=1}^{\infty}(-1)^k\sigma(m, k) = 0$ for $m \geq 2$. Since a permutation of $[m]$ with k cycles has sign $(-1)^{m+k}$, this is equivalent to the fact that for $m \geq 2$ the symmetric group S_m has equal numbers of odd and even permutations.

To prove (b), we show that $k!\sigma(m, k)$ satisfies the same recurrence as $(m-1)!b(k, m)$. The quantity $k!\sigma(m, k)$ is the number of ordered lists of the cycles comprising a permutation of m with k cycles. If $k = 1$, there are $(m-1)!$ of these, which equals $(m-1)!b(1, m)$. If $k \geq 1$, we count these lists by the number j of elements not in the first cycle. By picking these elements, forming an ordered list of k cycles for a permutation of them, and permuting the elements of the first cycle, we have

$$(k+1)!\sigma(m, k+1) = \sum_{j=k}^{m-1} \binom{m}{j} k!\sigma(j, k)(m-j-1)!.$$

Since $1/j + 1/(m-j) = m/(j(m-j))$, the numbers $(m-1)!b(k, m)$ satisfy the same recurrence.

Solved also by D. Callan, K. S. Kedlaya (student), P. Tracy, National Security Agency Problems Group, and the proposer.

Saturated Chain Partitions and Fibonacci Numbers

10199 [1992, 162]. *Proposed by Richard Stanley, Massachusetts Institute of Technology, Cambridge, MA.*

Given a finite partially ordered set P , let $f(P)$ denote the number of ways to partition the elements of P into pairwise disjoint nonempty saturated chains.

(a) Prove that if P_n is the product of two n -element chains, i.e., if $P_n = \{(i, j): 1 \leq i \leq n, 1 \leq j \leq n\}$, with $(i, j) \leq (i', j')$ if and only if $i \leq i'$ and $j \leq j'$, then $f(P_n) = \prod_{j=1}^n F_{2j}^2$, where F_k is the k th Fibonacci number.

(b) If every element of P covers at most two elements and is covered by at most two elements, prove that $f(P)$ factors into Fibonacci and Lucas numbers.

Solution by William Y. C. Chen, Los Alamos National Laboratory, Los Alamos, NM. We first define an operation on a poset P that preserves $f(P)$. Given an element $x \in P$, let P' be the poset obtained by replacing x by two incomparable elements x', x'' , one of which is a minimal element covered by the elements covering x in P , the other of which is a maximal element covering the elements covered by x in P . A partition of P becomes a partition of P' by splitting the chain containing x , and this map is reversible, so $f(P) = f(P')$. We call this operation *splitting* x . Let P^* be the poset obtained from P by splitting every element that is not maximal or minimal. It suffices to compute $f(P^*)$.

(a) For P_n , the poset P_n^* is a disjoint union of *zigzags*, where the zigzag Z_k is the poset with all elements either maximal or minimal whose comparability graph is a k -vertex path. By considering the length of the chain containing a given endpoint of this comparability graph, we find that $f(Z_k) = f(Z_{k-1}) = f(Z_{k-2})$ for $k > 2$. Since $f(Z_1) = 1$ and $f(Z_2) = 2$, we have $f(Z_k) = F_{k+1}$. The poset P_n^* is the disjoint union of $Z_3, Z_5, \dots, Z_{2n-1}, Z_{2n-1}, Z_{2n-3}, \dots, Z_3$; one zigzag for each consecutive pair of levels. Since $F_2 = 1$, we have $f(P_n) = \prod_{j=1}^n F_{2j}^2$.

(b) When every element of P covers at most two elements and is covered by at most two elements, P^* consists of disjoint zigzags and *crowns*, which are the posets whose comparability graphs are even cycles. Let Q_{2k} denote the crown with $2k$ elements. By considering whether a fixed edge of this comparability graph is a chain in a partition, we find that $f(Q_{2k}) = f(Z_{2k}) + f(Z_{2k-2}) = F_{2k+1} + F_{2k-1}$. Since $L_1 = 1$, $L_2 = 3$, and $L_n = L_{n-1} + L_{n-2}$ for $n > 2$, it is immediate by induction that the n th Lucas number L_n satisfies $L_n = F_{n+1} + F_{n-1}$. Thus we have $f(Q_{2k}) = L_{2k}$, which completes the proof.

Solved also by R. J. Chapman (U. K.), R. Stong, and the proposer. Part (a) only solved by J. C. Binz (Switzerland), and the Anchorage Math Solutions Group.

Still Uniquely Fibonacci

10203 [1992, 265]. *Proposed by Ivan Vidav, University of Ljubljana, Ljubljana, Yugoslavia.*

Suppose that a, b, c and d are positive integers satisfying the two relations

$$b^2 + 1 = ac \quad \text{and} \quad c^2 + 1 = bd.$$

Prove that $a = 3b - c$ and $d = 3c - b$.

Solution by Nasha Komanda, Central Michigan University, Mt. Pleasant, MI. Note the equivalence of $a = 3b - c$ to $b^2 + 1 = (3b - c)c$ to $b^2 + c^2 + 1 = 3bc$ to $d = 3c - b$. It therefore suffices to prove that if c divides $b^2 + 1$ and b divides $c^2 + 1$, then $b^2 + c^2 + 1 = 3bc$. We may assume $b \leq c$ and prove this assertion by induction on $b + c$.

If $b + c = 2$, then $b = c = 1$ and the assertion holds. If $c > 1$, then $b \neq c$, which implies $b \leq c - 1$, then $b^2 + 1 < c^2$, and $a < c$. The equalities $b^2 + 1 = ac$ and $c^2 + 1 = bd$ imply that $(b^2 + 1)^2 = a^2(bd - 1)$ and therefore $1 \equiv -a^2 \pmod{b}$, which means b divides $a^2 + 1$. Since a divides $b^2 + 1$ and $a + b < b + c$, we can apply the induction hypothesis to conclude that $a^2 + b^2 + 1 = 3ab$.

Now, $(b^2 + 1)^2 = a^2c^2 = (3ab - b^2 - 1)c^2 = 3abc^2 - (b^2 + 1)c^2$. Therefore, $(b^2 + 1)(b^2 + c^2 + 1) = 3abc^2 = 3bc(b^2 + 1)$. We conclude $b^2 + c^2 + 1 = 3bc$, which completes the proof.

Editorial comment. Several solvers noted that the solution to E3210 [1987, 457; 1988, 879] showed that if $b < c$, $b^2 \equiv -1 \pmod{c}$, and $c^2 \equiv -1 \pmod{b}$, then $(b, c) = (F_{2k-1}, F_{2k+1})$, where the F_n are Fibonacci numbers. It then follows quickly that $b^2 + c^2 + 1 = 3bc$ and that b and c satisfy the conditions of this problem.

In addition to references provided in the solution to E3210, Gerry Meyerson mentioned W. Sierpinski and A. Schinzel, Sur l'équation $x^2 + y^2 + 1 = xyz$, *Matematiche, Catania* 10(1955), 30-36; MR 17 711e.

Some solvers taking a more general approach used the fact that the continued fraction expansion of $(k - 2 + \sqrt{k^2 - 4})/2$ is purely periodic with period $(k - 2, 1)$ for all integers $k > 2$.

Robert J. Weber mentioned some related results:

- (1) $x^2 + y^2 = k(xy + 1)$ if and only if $k = l^2$,
- (2) $x^2 + y^2 + 1 = k(xy + 1)$ if and only if $k = l^2 + 1$,
- (3) $x^2 + y^2 = k(xy - 1)$ if and only if $k = 5$, and
- (4) $x^2 + y^2 - 1 = kxy$ has infinitely many such solutions for every $k \geq 2$.

He, along with A. Sudbery and the IMO 1992 Indian Team, noted that (1) was a problem in the 1988 International Mathematical Olympiad.

Based on the similarity in method of solution between the Olympiad problem and the present one, A. Sudbery suggested that there might be a general proposition which would unify these results. The additional references indicate that some work has already been done in this direction. A distillation of this brew may prove useful.

Solved by 53 readers and the proposer.

Square Functions

10235 [1992, 571]. *Proposed by Daniel Goffinet, Saint Étienne, France.*

(a) Determine the set \mathcal{F} of those continuous maps f from \mathbb{R}^2 to \mathbb{R} such that, for every rectangle $ABCD$, one has $f(A) + f(C) = f(B) + f(D)$.

(b) Let $KLMN$ be a quadrangle in the plane such that $f(K) + f(M) = f(L) + f(N)$ for every $f \in \mathcal{F}$. Is it true that $KLMN$ must be a rectangle?

Solution by Robert B. Israel, University of British Columbia, Vancouver, B. C., Canada. Part (a): \mathcal{F} consists of all functions of the form $f(\mathbf{x}) = a + \mathbf{b} \cdot \mathbf{x} + c\mathbf{x} \cdot \mathbf{x}$, where a and c are real and $\mathbf{b} = (b_1, b_2) \in \mathbb{R}^2$. Moreover, this is true even if “rectangle” is replaced by “square”.

Proof: First note that all functions of the above form belong to \mathcal{F} . For if $KLMN$ is a rectangle, write $L = K + \mathbf{u}$, $N = K + \mathbf{v}$, $M = K + \mathbf{u} + \mathbf{v}$ where \mathbf{u} and \mathbf{v} are orthogonal vectors, and we have $F(K) + F(M) = F(L) + F(N) = 2a + \mathbf{b} \cdot (2K + \mathbf{u} + \mathbf{v}) + c(2K \cdot K + K \cdot \mathbf{u} + K \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v})$.

Now suppose $f \in \mathcal{F}$. Let $r > 0$ be given. We can uniquely determine a , \mathbf{b} and c so that the function $g(\mathbf{x}) = a + \mathbf{b} \cdot \mathbf{x} + c\mathbf{x} \cdot \mathbf{x}$ agrees with $f(\mathbf{x})$ on the four points $(0, 0)$, $(r, 0)$, $(-r, 0)$ and $(0, r)$, namely

$$\begin{aligned} a &= f(0, 0) \\ b_1 &= \frac{f(r, 0) - f(-r, 0)}{2r} \\ b_2 &= \frac{2f(0, r) - f(r, 0) - f(-r, 0)}{2r} \\ c &= \frac{f(r, 0) + f(-r, 0) - 2f(0, 0)}{2r^2}. \end{aligned}$$

Note that since both f and g satisfy the given identity, if they agree at three vertices of a square then they agree at the fourth vertex. Now I claim that g agrees with f at all points of the square lattice $r\mathbb{Z}^2$. First we obtain agreement at $(0, -r)$ by using the square $(-r, 0)(0, r)(r, 0)(0, -r)$. Next, we use mathematical induction to show that for every positive integer N , there is agreement at all lattice points with $|x| + |y| \leq Nr$. We already have the case $N = 1$. Given that it is true for N , we obtain agreement at $(mr, (N + 1 - m)r)$ with $1 \leq m \leq N$ by using the squares $((m - 1)r, (N + 1 - m)r), (mr, (N + 1 - m)r), (mr, (N - m)r), ((m - 1)r, (N - m)r)$. Reflecting these squares about the x and/or y axes, we obtain also the points $(\pm mr, \pm (N + 1 - m)r)$. We obtain $(0, (N + 1)r)$ by using the

square $(0, (N+1)r), (r, Nr), (0, (N-1)r), (-r, Nr)$, and similarly we obtain $(0, -(N+1)r)$ and $(\pm(N+1)r, 0)$.

The definition of g appeared to depend on r . In fact, however, every rational r will give the same g . For if r_1 is an integer multiple of r_2 , $r_1\mathbb{Z}^2 \subseteq r_2\mathbb{Z}^2$ so the g defined for r_1 must agree with the g defined for r_2 . Since any two positive rationals are integer multiples of a single positive rational, they must have the same g .

Thus we have that f and g agree on all points with rational coordinates. By continuity, they agree everywhere.

Part (b): Yes, it must be a rectangle.

Considering first a function of the form $f(\mathbf{x}) = \mathbf{b} \cdot \mathbf{x}$, we have $\mathbf{b} \cdot (K - N) = \mathbf{b} \cdot (L - M)$. Since this is true for all \mathbf{b} , $K - N = L - M$, i.e., $KLMN$ is a parallelogram. Thus we can write $L = K + \mathbf{u}$, $N = K + \mathbf{v}$, $M = K + \mathbf{u} + \mathbf{v}$. Now using the function $f(\mathbf{x}) = (\mathbf{x} - K) \cdot (\mathbf{x} - K)$, we have $0 = f(K) + f(M) - f(L) - f(N) = 2\mathbf{u} \cdot \mathbf{v}$. Thus \mathbf{u} and \mathbf{v} are orthogonal, i.e., $KLMN$ is a rectangle.

Editorial comment. Vu Ha Van showed that one obtains the same solution of (a) under the assumption that the functions in \mathcal{F} to be Lebesgue measurable rather than continuous. However, using non-measurable functions $g: \mathbb{R} \rightarrow \mathbb{R}$ satisfying $g(x+y) = g(x) + g(y)$, it is easy to produce additional functions $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ satisfying $f(A) + f(C) = f(B) + f(D)$ for every parallelogram.

Solved also by F. Alouges & R. Cerf (France), R. J. Chapman (U.K.), M. Dindos (Slovakia), C. P. Grant, M. Hejny (Slovakia), H. H. Johnson, S. Kanetkar, I. Kastanas, P. Kinion, N. Komanda, O. P. Lossers (The Netherlands), K. Merryfield, M. Mócsy (Hungary), H. V. Vu (student, Hungary), University of Wyoming Problem Circle, and the proposer. Four incorrect solutions were also received.

Conservative Paths in Continuous Vector Fields

10239 [1992, 674]. *Proposed by Ismor Fischer, Naval Postgraduate School, Monterey, CA.*

A continuous vector field \vec{F} (in \mathbb{R}^2 or \mathbb{R}^3) and a simple closed curve Γ are given. Show that, for every point $x \in \Gamma$, there exists a point $y \in \Gamma$ and a path γ from x to y (nontrivial if $x = y$) such that the work $W = \int_{\gamma} \vec{F} \cdot d\vec{r}$ is zero.

Solution by Frédéric Brulois, California State University—Dominguez Hills, Carson, CA. We prove the stronger (and more simply stated)

Proposition. *Let U be an open set of \mathbb{R}^n , $x \in U$, and $\mathbf{F}: U \rightarrow \mathbb{R}^n$ a continuous vector field on U . Then there exists a nontrivial (real analytic) path $\gamma: [0, 1] \rightarrow U$ from x back to x such that $\int_{\gamma} \mathbf{F} \cdot d\mathbf{r} = 0$.*

Proof: Let C be a circle through x which is small enough that the convex hull of C is contained in U . Let $\beta: \mathbb{R} \rightarrow U$ be a constant speed parameterizations of C such that $\beta(0) = \beta(1) = x$ and $\beta(t) \neq x$ for $0 < t < 1$. For each $s \in [0, 1]$, define $\alpha_s: [0, 1] \rightarrow \mathbb{R}^n$ by setting

$$\alpha_s(t) = (1-s)\beta(t) + s\beta(-2t), \quad \text{for } 0 \leq t \leq 1.$$

Clearly $\alpha_0 = \beta$, while α_1 maps twice around the circle C in the direction opposite that of β . Also α_s maps into the convex hull of C , hence into U . It is easily seen that each α_s defines a nontrivial path. Finally α_s is a real analytic map since β is.

Now consider the function

$$f(s) = \int_{\alpha_s} \mathbf{F} \cdot d\mathbf{r} = \int_0^1 \mathbf{F}(\alpha_s(t)) \cdot \alpha'_s(t) dt, \quad \text{for } 0 \leq s \leq 1.$$

The function $(s, t) \mapsto \mathbf{F}(\alpha_s(t)) \cdot \alpha'_s(t)$ is continuous, hence uniformly continuous on the compact set $[0, 1] \times [0, 1]$. Thus f is continuous on $[0, 1]$. Now

$$f(0) = \int_{\beta} \mathbf{F} \cdot d\mathbf{r} \quad \text{and} \quad f(1) = \int_{-\beta \cdot 2} \mathbf{F} \cdot d\mathbf{r} = -2 \int_{\beta} \mathbf{F} \cdot d\mathbf{r}.$$

Therefore, by the Intermediate Value Theorem, there exists $s_0 \in [0, 1]$ such that $f(s_0) = 0$. Take $\gamma = \alpha_{s_0}$. Then $\int_{\gamma} \mathbf{F} \cdot d\mathbf{r} = f(s_0) = 0$.

The proposition shows that we may always take $y = x$: the curve γ in the statement of the problem turns out to be irrelevant! The phenomenon is purely local.

Editorial comment. It would be tempting to use of $\beta(-t)$ instead of $\beta(-2t)$ in the definition of α_s . However, in that case $\alpha_{1/2}$ would consist of a curve which moves along a line segment once in each direction. Such a curve should be considered as a *trivial* path from x to itself, since the work along such a path is zero for all vector fields.

Solved also by I. Kastanas, O. P. Lossers (The Netherlands), A. Nijenhuis, Western Maryland College Problems group, and the proposer.

Collaborating editors: David F. Appleyard, Paul T. Bateman, Duane M. Broline, Barry W. Brunson, Frank S. Cater, Gulbank D. Chakerian, Underwood Dudley, Gerald A. Edgar, Michael A. Filaseta, Ira M. Gessel, Richard A. Gibbs, Jerrold R. Griggs, Douglas A. Hensley, John R. Isbell, Mourad E. H. Ismail, Murray Klamkin, Daniel J. Kleitman, Frederick W. Luttmann, Frank B. Miles, Richard Pfiefer, Stephen L. Portnoy, J. O. Shallit, John Henry Steelman, Kenneth B. Stolarsky, David E. Tepper, Douglas B. Tyler, Daniel Ullman, and William E. Watkins.

“Mathematics was born and nurtured in a cultural environment. Without the perspective which the cultural background affords, a proper appreciation of the content and state of present-day mathematics is hardly possible.”

—R. L. Wilder

REVIEWS

Edited by **Darrell Haile**
Indiana University, Bloomington IN 47405

Revolutions in Mathematics. Edited by Donald Gillies, Oxford University Press, New York, 1992, viii + 353, \$98.00

Reviewed by **Michael S. Mahoney**

The theme of this book may need explication for potential readers who are not historians or philosophers of science. The title should have a question mark. At issue among the authors is precisely whether, and in what sense, revolutions occur in mathematics. That question hangs on what one means by "revolution," and behind that lies the larger question of the extent to which the history of mathematics is amenable to the historiographical approaches connoted by the notion of revolutions in science.

The touchstone is Thomas S. Kuhn's model of scientific change as set forth in *The Structure of Scientific Revolutions* in 1962, amended in a postscript added to the work in 1970, and subsequently articulated by a host of commentators, both friendly and not. In *Structure* Kuhn argued that science is inherently conservative. It consists normally of puzzle-solving: finding answers to problems well defined by a unanimously accepted *paradigm* (or, in a later version, *disciplinary matrix*) that dictates how they should be solved and that offers assurance that they have solutions, however difficult or recalcitrant they might appear. Practice departs from the norm when puzzles resist efforts at solution by ever more skillful practitioners. Unsolved problems may be tolerated for a time as *anomalies*, or problems that don't fit the rules, but at some point, a solution appears to some practitioners as crucial to the continued viability of the field. Either it is resolved, or some fundamental change in the rules is required to make it tractable. The paradigm has reached a *crisis*, the point of decision whether it will survive. Either the problem succumbs, or the paradigm does. Where a solution demands a new paradigm, or rather itself constitutes a new paradigm, a revolution occurs, marking a radical discontinuity in the development of the scientific field in question.

The discontinuity takes both intellectual and social form. The old paradigm may lose meaning entirely, as phlogiston did when replaced by the oxygen theory of combustion. But, even where portions of the old paradigm are incorporated in the new, they acquire new or altered meaning through a change in the basic concepts of the field, as in the shift from classical to relativistic mechanics. Practitioners of the old paradigm find their skills obsolete, their expertise no longer of interest. That is where scientific revolutions become as bloody as intellectual combat is likely to get. Reputations are at stake, as is the work of whole careers, and the people involved understandably resist losing them.

Kuhn's model appealed to historians because it opened scientific change to operative factors external to the science at issue. How problems attract attention, how they become critical, whence new paradigms arise, how the new is compared to the old: none of these questions can be addressed, much less answered, from

within either the old or the new paradigm. Revolutions signal a break in the conceptual autonomy that seems to insulate normal science from influences beyond its own rules. Through revolutions, historians could root science in place and time, in social structures and cultural contexts.

Sensitive to the new historiography of science of which Kuhn was both symptom and stimulus, Michael Crowe in 1974 presented to a colloquium on the history of modern mathematics at the American Academy of Arts and Sciences his “Ten ‘laws’ concerning patterns of change in the history of mathematics,” which appeared in *Historia Mathematica* the following year and forms Chapter 1 of the present work. The tenth law asserts that “Revolutions never occur in mathematics.” Revolutions require that “some previously existing entity (be it king, constitution, or theory) must be overthrown and irrevocably discarded. [19]” But while that may occur in mathematical symbolism, terminology, methodology, or even historiography, it does not happen *in* mathematics itself. The substance of the old retains its validity, albeit in a different form, while the new takes a place alongside it.

Crowe’s assertion drew two major responses, republished here with the authors’ afterthoughts (Chapters 2–5). In “T. S. Kuhn’s theories and mathematics,” first published in 1976, Herbert Mehrtens disagreed with Crowe’s separation of substance from form in mathematics: “. . . there are events in the history of mathematics that might be termed ‘revolutions’, and . . . there is no point in distinguishing these events with respect to their being ‘in’ mathematics or somewhere else. [26]” Yet, Mehrtens quickly added that the term “revolution” had more emotive than analytic force as an historiographical concept, since the implicit political analogy was difficult to sustain. “If there is to be any serious use of the metaphor,” “he reiterates in his 1992 appendix, “then it should aim at the structures of power and legitimacy before and after the event. In mathematics and the sciences one has to press the metaphor hard to make it work in this sense. [43]” Kuhn’s notions of “normal science” and “anomaly” go a long way in explaining Crowe’s ten laws. Mehrtens argues, but “crisis” and hence “revolution” have limited explanatory power. In “Conceptual revolutions and the history of mathematics: two studies in the growth of knowledge,” read as a paper in 1974 and expanded for publication in 1984, Joseph Dauben acknowledged the conservative, or cumulative, nature of mathematics, while at the same time arguing that episodes such as the Greek discovery of incommensurable magnitudes and Cantor’s theory of transfinite sets transgressed the accepted limits of mathematical thought at the time, causing a reformulation of what had gone before and provoking the sort of resistance to change that Kuhn had signaled as a concomitant of revolution. His appendix adds two more episodes to the list: Cauchy’s rigorous foundation of the calculus and Robinson’s non-standard analysis.

The remaining chapters, specially commissioned for this volume, take up the issue from there, addressing what might be called “Crowe’s quandary.” Clearly, there have been radical changes in mathematical thought and practice, changes that one would want to call “revolutionary” in the common usage of the term: the calculus, non-Euclidean geometry, or abstract algebra, to name a few. Clearly, too, they have left older mathematics in place. The school curriculum bears witness to the continuing validity of arithmetic, geometry, and algebra as they were practiced five hundred years ago and as they today continue to serve as general foundation for the subject. (Interestingly, none of the contributors notes the pedagogical issues implicit in the notion of revolutions, namely, how students shift their thinking as the curriculum recapitulates these episodes of conceptual and epistemological upheaval.)

How then does mathematics change while remaining the same? What changes deserve to be called “revolutionary”? By what definition or model of “revolution”? Very little consensus emerges from the nine contributors. For the most part, they share a sense of revolution as a radical break with the past, a point at which new ways of thinking become possible, but they have little in common beyond that. In several instances, some see revolutions where others do not. Where they agree on what episodes were revolutions, in particular, non-Euclidean geometry, they differ on what made them revolutionary.

Paolo Mancosu’s detailed exposition (Chapter 6) of Descartes’s analytic geometry leads the reader at first to think that something revolutionary was going on, but, after reviewing both pre-Kuhnian and post-Kuhnian claims to that effect, Mancosu has his doubts. His scepticism is tempered by his acknowledged disregard of Descartes’s theory of equations and of arguments that Descartes’s new geometry was part of a larger, revolutionary shift from geometric to algebraic modes of thinking. Emily Grosholz (Chapter 7) makes more of that shift, as algebra provided the mediating element in Leibniz’s unification of geometry and number theory to form his calculus, which in turn provided a vehicle for uniting mechanics with mathematics. In thus opening new domains along the borders of its constituents, the calculus also transformed those older fields, thus giving it its revolutionary impact. But why it is particularly revolutionary, rather than simply creative or innovative, is hard to discern from Grosholz’ focus on Leibniz, which leaves out of the picture contemporaries pursuing the same ends in the same and different ways.

Giulio Giorello (Chapter 8) mixes Kuhn’s model with the political metaphor to perceive in the eighteenth-century debate over fluxions a reflection of England’s “Glorious Revolution” of 1689, which retained the monarchy by reconciling it with the new forms of political legitimacy embodied in Parliament. Here Newton’s fluxions as he presented them challenged the established “paradigm of legitimacy”, which Berkeley sought to defend in a “counter-revolution”. Maclaurin saved Newton’s revolution by reconciling the new modes of reasoning with traditional standards of geometric rigor.

Following Crowe’s lead, Caroline Dunsmore (Chapter 11) allows revolutions only at the level of the “metamathematical values of the community that define the telos and methods of the subject, and encapsulate general beliefs about its nature” and not at the “object-level” of “concepts, terminology and notation, definitions, axioms, and theorems, methods of proof and problem-solutions, and problems and conjectures. [211]” By her criteria, the discovery of incommensurables, negative and imaginary numbers, non-commutative algebra, and transfinite sets constitute revolutions, while symbolic algebra and the calculus do not. By contrast, looking at algebraic number theory and non-Euclidean and projective geometry in the nineteenth century, Jeremy Gray (Chapter 12) argues that “although the objects of study remained superficially the same, the way they were defined, analysed theoretically, and thought about intuitively was entirely transformed” [245], constituting a new framework of mathematics incompatible with its predecessor.

Two contributors come at the question from altogether different directions. Examining the emergence of non-Euclidean geometry, Yuxing Zheng (Chapter 9) sees two sorts of innovation in mathematics: that which runs in the same direction as current thinking and that which runs in the opposite direction. The latter provokes crisis by introducing contradictions within mathematics, and revolution consists in the restoration of harmony by the expansion of concepts to transform the contradictions into alternatives. Zheng calls this the “harmonious principle of

counter-way thinking". Considering "The geometrical vision of space" (Chapter 10), Luciano Boi resists any notion of "revolution" that applies sociological or historical categories to the development of mathematics. Rather, a wholly internal process of "hermeneutics" leads to two forms of growth: "an indefinite 'rewriting' of mathematical problems in the framework of one particular tradition or in different traditions" and "a radical change in the way a mathematical concept or field is conceived" [200–201]. In the latter case, the measure of the "revolutionary" nature of that change is the capacity of the new concept or field to provide fruitful models for the other fields, thus embracing what has gone before while at the same time rewriting or reconstruing it.

The volume's editor, Donald Gillies, has the next to last word before turning the discussion back to Michael Crowe for his afterthoughts. In arguing for "The Fregean revolution in logic", Gillies seeks to reconcile radical change with preservation of the old order by contrasting the "Russian" model of revolution with the "Franco-British". The former discarded the monarchy, the latter retained or restored it "with greatly diminished powers". While science has had its Russian revolutions, mathematics is restricted to the Franco-British sort. Gillies's candidate for revolution tests that limitation, pointing to the all but complete displacement of Aristotelian syllogistic logic by Frege's propositional and first-order predicate calculus, reinforced by Peano's determination of previously undetected exceptions to some syllogisms. In undermining of uniqueness of Aristotle's logic, the new system had metalevel implications similar to those of non-Euclidean geometry and for that reason provoked resistance from several quarters, thus fitting both Dunsmore's and Dauben's models of revolutions.

According to Garrett Birkhoff, his father's initial response to lattice theory was to ask "what one could do with lattices that couldn't be done without them" (*Papers on Algebra and Topology*, 577–8). One may ask a similar question of *Revolutions in Mathematics*. What does the notion of "revolutions" do for history of mathematics that can't be done without it? Mehrtens alone raises the issue and concludes, "Nothing". As the other authors debate the possibility of revolutions, testing their arguments against a series of well known episodes, it never becomes clear just what is at stake historiographically, despite Gillies's admirable introductory review of the different positions. Grosholz unintentionally provides a hint when she concedes at the start of her essay that "the term 'revolutionary' has come to have an honorific sense in the philosophy of science, and I would like to see Leibniz properly honoured." [116] It sounds like a throw-away line, but it captures a tacitly shared assumption among most of the authors, with the obvious exception of Mehrtens, who observes in his "Appendix" (1992) that "Revolutions are no longer what they were. [43]". All of the developments proposed as revolutions here are "good" revolutions. They aim toward mathematics as it is currently conceived and thus, perhaps ironically, have a cumulative effect. To read through *Revolutions in Mathematics* is to watch the modern curriculum develop one revolutionary step at a time. None of the revolutions leads in the wrong direction. There are no failed revolutions, nor is there significant opposition, reaction, or counter-revolution. Giorello casts Berkeley in that last role, but only metaphorically and ultimately in ways supportive of the new calculus. Only Herbert Breger puts opposition in the spotlight with his study of Paul Finsler's intuitive set theory (Chapter 13), "a restoration that failed", less because of its inconsistency than because of its lack of fit with the newly dominant style of formalist mathematics espoused by Hilbert.

Historians of political revolutions take the opposition more seriously. For them, revolutions may succeed or fail, and successive revolutions—and restorations and counter-revolutions—may take a society in different, even opposite, directions. The present is neither explanation nor criterion of revolutions in the past. They are a matter of choices among alternatives at the time, and all choices equally require explanation. The notion of revolutions in mathematics might acquire analytic force if historians paid more attention than they have to those who resisted the transformations that constitute the present-day subject. Studying “reaction” as well as “revolution” would bring out the factors of stasis that define change and that show up, in mathematics especially, as continuations of the “old order”, thus producing Crowe’s quandary. As the contributions to this volume show, whether or not we call these formative episodes “revolutions”, they ultimately involve clashes of intellectual and social values and hence bring into sharp focus how those values shape mathematics. Viewing them in that light means taking the values of both sides seriously and bearing in mind that at any juncture things could have gone otherwise and still have been mathematics.

History of Science Department
Princeton University
Princeton, NJ 08544

“Regarding all these basic topics in infinitesimal calculus which we teach today as canonical requisites . . . , the question is never raised, ‘Why so?’ or ‘How does one arrive at them?’ Yet all these matters must at one time have been goals of an urgent quest, answers to burning questions, at the time, namely, when they were created. If we were to go back to the origins of these ideas, they would lose that dead appearance of cut-and-dried facts instead take on fresh and vibrant life again.”

—*O. Toeplitz*

TELEGRAPHIC REVIEWS

Edited by **Arnold Ostebee and Paul Zorn**

with the assistance of the Mathematics Departments of
Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T : Textbook	P : Professional Reading	1-4 : Semester
C : Computer Software	L : Undergraduate Library	** : Special Emphasis
S : Supplementary Reading	13 : Grade Level	?? : Questionable

Readers are advised that price information is subject to change. Selected books and software packages receive a second, more extensive review in the *Monthly*.

Books and software submitted for review should be sent to *Book Reviews Editor*, *American Mathematical Monthly*, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098.

General, S*(16-17), C. *Introduction to Maple*. André Heck. Springer-Verlag, 1993, xiii + 497 pp, \$39. [ISBN 0-387-97662-0] An excellent introduction to the computer algebra/graphics system *Maple*. Assumes little previous experience with *Maple*, but moves rapidly to use of *Maple* to solve complex problems. Several in-depth examples worked in detail. Shows nicely not only how *Maple* works, but why it does what it does. MPR

General, P. *I.M. Gelfand Seminar*. Eds: Sergei Gelfand, Simon Gindikin. Adv. in Soviet Math., V. 16. AMS, 1993. *Part 1*, xiv + 241 pp, \$71 [ISBN 0-8218-4118-1]; *Part 2*, xiv + 208 pp, \$65. [ISBN 0-8218-4119-X] 14 invited papers honoring Gelfand's 80th birthday.

General, P. *Proceedings of the St. Petersburg Mathematical Society, Volume I*. Eds: O.A. Ladyzhenskaya, A.M. Vershik. Transl. Ser. 2, V. 155. AMS, 1993, xiii + 223 pp, \$99. [ISBN 0-8218-7505-1] 8 papers in diverse mathematical areas.

General, S(13). *Mathematik-Vorkurs: Übungsbuch und Arbeitsbuch für Studienanfänger*. Wolfgang Schäfer, Kurt Georgi. BG Teubner Leipzig, 1993, 470 pp, DM 44 (P). [ISBN 3-8154-2038-5] A compendium, for independent study, of mathematics taught in German secondary schools. Algebra, trigonometry, logic, analytic geometry, vectors, calculus. Many exercises, with solutions. JD-B

General, P. *Mathematical T_EX by Example*. Arvind Borde. Academic Pr, 1993, xi + 352 pp, \$29.95 (P). [ISBN 0-12-117645-2] Shows how to use T_EX to typeset mathematics. Input

commands and typeset output appear on facing pages. AO

General, P. *Transactions of the Tenth Army Conference on Applied Mathematics and Computing*. US Army Research Office (POB 12211, Research Triangle Park, NC 27709-2211), 1993, xxi + 770 pp, (P). 42 papers on topics in natural language processing, wavelet analysis, variational methods, small sample asymptotics, computational fluid dynamics, control theory, parallel programming, computational algebraic geometry, and mathematical aspects of materials science, etc.

General, P. *Mathematical Computation with Maple V: Ideas and Applications*. Ed: Thomas Lee. Birkhäuser, 1993, viii + 199 pp, \$34.50 (P). [ISBN 0-8176-3724-9] Proceedings of the 1993 Maple Summer Workshop and Symposium at the University of Michigan.

General, P. *An Introductory Guide to Scientific Visualization*. R.A. Earnshaw, N. Wiseman. Springer-Verlag, 1992, xvi + 156 pp, \$49. [ISBN 0-387-54664-2] Written for novices. First part: basic ideas, vocabulary, and applications. Second part: descriptions of many commercial and public domain systems. AO

Recreational Mathematics, L*. *The Wohascum County Problem Book*. George T. Gilbert, Mark I. Krusemeyer, Loren C. Larson. Dolciani Math. Expos., No. 14. MAA, 1993, ix + 233 pp, \$26 (P). [ISBN 0-88385-316-7] 130 challenging, original problems with clear, elegant solutions (often several per problem). A few problems use basic linear or abstract algebra; most require nothing beyond calculus;

many are accessible to high school students. A must for any problem solver. DH

Precalculus, T(13: 1). *College Algebra and Trigonometry, Third Edition.* Michael Sullivan. Dellen, 1993, xxi + 1034 pp. [ISBN 0-02-418305-9] New in this edition: more emphasis on visualization; exercises and examples using graphing calculators; discussion/writing questions; some revised and/or reorganized material. AO

Precalculus, S(13). *Schaum's Outline of Theory and Problems of Mathematical Methods for Business and Economics.* Edward T. Dowling. McGraw-Hill, 1993, ix + 384 pp, \$12.95 (P). [ISBN 0-07-017697-3] Includes topics from high school algebra, linear algebra, linear programming, calculus. 1742 problems, most with complete solutions. DH

Education, P, L. *Developments in School Mathematics Education Around the World.* Eds: Izaak Wirszup, Robert Streit. Univ. of Chicago School Math. Proj. NCTM, 1992, xv + 501 pp, \$22 (P). [ISBN 0-87353-356-9] Proceedings of the Third UCSMP International Conference on Mathematics Education (Chicago, 1991).

Education, S(14-17). *Making Sense of Data.* Mary M. Lindquist, *et al.* Addenda Ser., Grades K-6. NCTM, 1992, viii + 48 pp, \$9.50 (P), [ISBN 0-87353-318-6]; *Number Sense and Operations.* Grace M. Burton, *et al.* Addenda Ser., Grades K-6. NCTM, 1993, viii + 55 pp, \$9.50 (P), [ISBN 0-87353-319-4]; *Patterns.* Terrence G. Coburn, *et al.* Addenda Ser., Grades K-6. NCTM, 1993, ix + 53 pp, \$9.50 (P), [ISBN 0-87353-320-8]; *Geometry and Spatial Sense.* John Del Grande, *et al.* Addenda Ser., Grades K-6. NCTM, 1993, viii + 55 pp, \$9.50 (P). [ISBN 0-87353-317-8] The contents of these four books duplicate the contents of the seven K-6 grade level books in the Addenda Series (TR, April 1993). Grade level books were divided into four topic areas; topic books are divided into seven grade levels. Topic books also include additional background information, extensive bibliographies. MW

Education, P. *Mathematicians and Education Reform, 1990-1991.* Eds: Naomi D. Fisher, Harvey B. Keynes, Philip D. Wagreich. CBMS Issues in Math. Educ., V. 3. AMS and MAA, 1993, x + 185 pp, \$62 (P). [ISBN 0-8218-3503-3] 14 articles on projects and issues in mathematics education reform.

Education, S, P*, L. *The Geometric Supposer: What Is It a Case Of?* Eds: Judah L. Schwartz, Michal Yerushalmy, Beth Wilson. Tech. in Educ. Ser. Lawrence Erlbaum Assoc, 1993, xi

+ 254 pp, \$49.95. [ISBN 0-8058-0720-9] Reports on use of the Geometric Supposer: problems in learning and teaching, how learning behavior changes. Authors include developers, math educators, teachers, and students. JNC

Education, P. *Numerical Cognition.* Ed: Stanislas Dehaene. Blackwell, 1993, 209 pp, \$18.95 (P). [ISBN 1-55786-444-6] 5 papers survey recent research.

History, P. *Élie Cartan (1869-1951).* M.A. Akivis, B.A. Rosenfeld. Transl. of Math. Mono., V. 123. AMS, 1993, xii + 317 pp, \$153. [ISBN 0-8218-4587-X] A scientific biography. Describes and evaluates Cartan's most important discoveries. AO

Logic, P. *Uncountably Categorical Theories.* Boris Zilber. Transl. of Math. Mono., V. 117. AMS, 1993, vi + 122 pp, \$97. [ISBN 0-8218-4586-1]

Logic, P. *Many-Valued Logics, Volume 1: Theoretical Foundations.* Leonard Bolc, Piotr Borowik. Springer-Verlag, 1992, xii + 292 pp, \$69. [ISBN 0-387-55926-4]

Graph Theory, P. *Graph Structure Theory.* Eds: Neil Robertson, Paul Seymour. Contemp. Math., V. 147. AMS, 1993, xiv + 688 pp, \$81 (P). [ISBN 0-8218-5160-8] Proceedings of the 1991 AMS-IMS-SIAM Joint Summer Research Conference on Graph Minors held at the University of Washington.

Graph Theory, T(16-17: 1), P, L. *The Petersen Graph.* D.A. Holton, J. Sheehan. Australian Math. Soc. Lect. Ser., V. 7. Cambridge Univ Pr, 1993, 353 pp, \$39.95 (P). [ISBN 0-521-43594-3] Studies areas of graph theory where the Petersen graph plays an important role (often as a counterexample). Covers basic definitions, coloring problems, sharks (a certain 4-edge-connected cubic graph), matchings, cases, hypohamiltonian graphs. Usable as a second course in graph theory for seniors (exercises included) or as reference for researchers (unsolved problems mentioned). LC

Discrete Mathematics, P. *Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms.* ACM and SIAM, 1993, 506 pp, (P). [ISBN 0-89871-313-7] 54 papers from the 1993 symposium in Austin, Texas.

Number Theory, P. *Number Theory With an Emphasis on the Markoff Spectrum.* Eds: Andrew D. Pollington, William Moran. Lect. Notes in Pure & Appl Math., V. 147. Marcel Dekker, 1993, viii + 321 pp, \$125 (P). [ISBN 0-8247-8902-4] 28 papers from the 1991 Conference on the Markoff Spectrum and Diophantine Approximation and Analytic Number Theory at Brigham Young University.

Number Theory, T*(17). *A Course in Computational Algebraic Number Theory.* Henri Cohen. Grad. Texts in Math., V. 138. Springer-Verlag, 1993, xxi + 534 pp, \$49. [ISBN 0-387-55640-0] Computer algorithms for calculating number-theoretic quantities. First six chapters cover fundamental ideas: Euclid's algorithm, solving equations modulo a prime, polynomial factoring, algebraic number field techniques, quadratic field techniques, computing Galois groups, etc. Chapter 7 introduces elliptic curves (including the Taniyama, Birch, and Swinnerton-Dyer Conjectures) and many important algorithms (e.g., for computing the rational points and the L -function of a curve). Chapters 8–10 treat integer factoring and primality testing. An impressive text—probably the standard for some time to come. MPR

Number Theory, P. *Local Fields and Their Extensions: A Constructive Approach.* I.B. Fesenko, S.V. Vostokov. Transl. of Math. Mono., V. 121. AMS, 1993, xv + 283 pp, \$118. [ISBN 0-8218-4613-2]

Number Theory, P. *Collected Papers of Kustaa Inkeri.* Eds: Tauno Metsänkylä, Paulo Ribenboim. Papers in Pure & Appl. Math., V. 91. Queen's Univ, 1992, xxxi + 566 pp, (P). [ISBN 0-88911-632-6]

Number Theory, P. *The Stickelberger Ideal in the Spirit of Kummer with Application to the First Case of Fermat's Last Theorem.* Vijay Jha. Queen's Papers in Pure & Appl. Math., No. 93. Queen's Univ, 1993, xiv + 181 pp, (P).

Group Theory, P. *Group Representations, Volume 2.* Gregory Karpilovsky. Math. Stud., V. 177. North-Holland (US Distr: Elsevier Science), 1993, xv + 902 pp, \$185.75. [ISBN 0-444-88726-1]

Group Theory, P. *The Admissible Dual of $GL(N)$ via Compact Open Subgroups.* Colin J. Bushnell, Philip C. Kutzko. Annals of Math. Stud., No. 129. Princeton Univ Pr, 1993, ix + 313 pp, \$59.50; \$24.95 (P). [ISBN 0-691-03256-4; 0-691-02114-7]

Group Theory, P. *Abelian Groups.* Eds: Laszlo Fuchs, Rüdiger Göbel. Lect. Notes in Pure & Appl. Math., V. 146. Marcel Dekker, 1993, xii + 260 pp, \$99.75 (P). [ISBN 0-8247-8901-6] Proceedings of the International Conference on Torsion-free Abelian Groups (Curaçao, 1991). 4 survey articles and 17 research articles.

Group Theory, P. *Arithmetic of Probability Distributions, and Characterization Problems on Abelian Groups.* G.M. Fel'dman. Transl. of Math. Mono., V. 116. AMS, 1993, 223 pp, \$127. [ISBN 0-8218-4593-4]

Group Theory, P. *Subgroups of Teichmüller Modular Groups.* Nikolai V. Ivanov. Transl. of Math. Mono., V. 115. AMS, 1992, xii + 127 pp, \$107. [ISBN 0-8218-4594-2]

Algebra, P. *Selected Papers in K -Theory.* Transl. Ser. 2, V. 154. AMS, 1992, ix + 195 pp, \$83. [ISBN 0-8218-7504-3]

Algebra, P. *Boolean Constructions in Universal Algebras.* A.G. Pinus. Math. & Its Applic., V. 242. Kluwer Academic, 1993, vii + 350 pp, \$149. [ISBN 0-7923-2117-0]

Algebra, P. *Ordered Algebraic Structures: The 1991 Conrad Conference.* Eds: J. Martinez, C. Holland. Kluwer Academic, 1993, xiii + 258 pp, \$110.50. [ISBN 0-7923-2258-4] 14 papers from a conference held at the University of Florida to honor Paul F. Conrad.

Algebra, P. *Kac Algebras and Duality of Locally Compact Groups.* Michel Enock, Jean-Marie Schwartz. Springer-Verlag, 1992, x + 257 pp, \$98. [ISBN 0-387-54745-2]

Algebra, P. *Theory of Commutative Fields.* Masayoshi Nagata. Transl: Masayoshi Nagata. Transl. of Math. Mono., V. 125. AMS, 1993, xv + 249 pp, \$125. [ISBN 0-8218-4572-1]

Algebra, P. *Algebras and Orders.* Eds: Ivo G. Rosenberg, Gert Sabidussi. NATO ASI Ser. C, V. 389. Kluwer Academic, 1993, xvii + 553 pp, \$220. [ISBN 0-7923-2143-X] 12 papers by invited speakers at the 1991 NATO Advanced Study Institute in Montréal.

Algebra, P. *Noncommutative Distributions: Unitary Representation of Gauge Groups and Algebras.* Sergio A. Albeverio, et al. Pure & Appl. Math., V. 175. Marcel Dekker, 1993, viii + 190 pp, \$99.75. [ISBN 0-8247-9131-2]

Calculus, S?(13). *Calculus Laboratories with Mathematica, Volume 1.* Michael G. Kerckhove, Van C. Nall. McGraw-Hill, 1993, 40 pp, (P). [ISBN 0-07-034220-2] 13 computer-based exercises (notebooks) with minimal Mathematica documentation; supplements standard Calculus I (some exercises refer to *Calculus and Analytic Geometry, Fifth Edition*, by Stein and Barcellos). JNC

Calculus, T(13: 3). *Calculus, Fifth Edition.* Stanley I. Grossman. Saunders College, 1992, xx + 1142 pp, \$68. [ISBN 0-03-096420-2] Multi-colored tome includes initial review chapter, early treatment of trigonometric functions; features more calculator use, more figures, ten new historical essays. (Third Edition, TR, January 1985.) JNC

Complex Analysis, P. *Theory of Entire and Meromorphic Functions: Deficient and Asymptotic Values and Singular Directions.* Zhang

- Guan-Hou. Transl. of Math. Mono., V. 122. AMS, 1993, xi + 375 pp, \$182. [ISBN 0-8218-4589-6]
- Complex Analysis, P.** *Carleman's Formulas in Complex Analysis: Theory and Applications.* Lev Aizenberg. Math. & Its Applic., V. 244. Kluwer Academic, 1993, xx + 299 pp, \$137. [ISBN 0-7923-2121-9]
- Complex Analysis, P.** *Entire and Subharmonic Functions.* Ed: B. Ya. Levin. Adv. in Soviet Math., V. 11. AMS, 1992, vii + 275 pp, \$147. [ISBN 0-8218-4110-6] 15 papers from the Research Seminar on the Theory of Functions at Kharkov University.
- Complex Analysis, P.** *The Cauchy Method of Residues, Volume 2: Theory and Applications.* Dragoslav S. Mitrinović, Jovan D. Kečkić. Math. & Its Applic., V. 259. Kluwer Academic, 1993, x + 191 pp, \$86.50. [ISBN 0-7923-2311-4]
- Differential Equations, P.** *Averaging in Stability Theory: A Study of Resonance Multi-Frequency Systems.* M.M. Hapaev. Math. & Its Applic., V. 79. Kluwer Academic, 1993, xiii + 279 pp, \$145. [ISBN 0-7923-1581-2]
- Differential Equations, P.** *Differential Inclusions in Nonsmooth Mechanical Problems: Shocks and Dry Friction.* Manuel D.P. Monteiro Marques. Progress in Nonlinear Diff. Eqts. & Their Applic., V. 9. Birkhäuser, 1993, x + 179 pp, \$69. [ISBN 0-8176-2900-9]
- Differential Equations, S*.** *Differential and Integral Equations through Practical Problems and Exercises.* Gheorghe Micula, Paraschiva Pavel. Texts in Math. Sci., V. 7. Kluwer Academic, 1992, ix + 395 pp, \$139. [ISBN 0-7923-1890-0] Hundreds of problems (with solutions) on first-order ODE's and PDE's, linear ODE's, Laplace transforms, integral equations, numerical methods. Nice resource. SK
- Differential Equations, P.** *Ordinary and Partial Differential Equations, Volume IV.* Eds: B.D. Sleeman, R.J. Jarvis. Pitman Res. Notes in Math. Ser., V. 289. Longman Scientific & Technical (US Distr: Wiley), 1993, 292 pp, \$41.95 (P). [ISBN 0-582-09137-3] Proceedings of 1992 University of Dundee conference.
- Differential Equations, P.** *Applications of Liapunov Methods in Stability.* A. Halanay, V. Rășvan. Math. & Its Applic., V. 245. Kluwer Academic, 1993, xi + 237 pp, \$115. [ISBN 0-7923-2120-0]
- Differential Equations, P.** *Asymptotic Properties of Solutions of Nonautonomous Ordinary Differential Equations.* I.T. Kiguradze, T.A. Chanturia. Math. & Its Applic., V. 89. Kluwer Academic, 1993, xiv + 331 pp, \$154. [ISBN 0-7923-2059-X]
- Differential Equations, P.** *Introduction to the General Theory of Singular Perturbations.* S.A. Lomov. Transl. of Math. Mono., V. 112. AMS, 1992, xviii + 375 pp, \$201. [ISBN 0-8218-4569-1]
- Differential Equations, P.** *Nevanlinna Theory and Complex Differential Equations.* Ilpo Laine. Stud. in Math., V. 15. Walter de Gruyter, 1993, viii + 341 pp, DM 154. [ISBN 3-11-013422-5]
- Partial Differential Equations, P.** *Hypo-Analytic Structures: Local Theory.* François Trèves. Math. Ser., V. 40. Princeton Univ Pr, 1992, xvii + 497 pp, \$65. [ISBN 0-691-08744-X]
- Partial Differential Equations, P.** *Nonlinear Stochastic Evolution Problems in Applied Sciences.* N. Bellomo, Z. Brzezniak, L.M. de Socio. Math. & Its Applic., V. 82. Kluwer Academic, 1992, xiv + 219 pp, \$93.50. [ISBN 0-7923-2042-5]
- Partial Differential Equations, P.** *Typical Singularities of Differential 1-Forms and Pfaffian Equations.* Michail Zhitomirskii. Transl. of Math. Mono., V. 113. AMS, 1992, xi + 176 pp, \$116. [ISBN 0-8218-4567-5]
- Partial Differential Equations, P.** *Developments in Partial Differential Equations and Applications to Mathematical Physics.* Eds: G. Buttazzo, G.P. Galdi, L. Zanghirati. Plenum Pr, 1992, viii + 246 pp, \$69.50. [ISBN 0-306-44311-2] Proceedings of a 1991 conference in Ferrara, Italy.
- Partial Differential Equations, P.** *Diffusion Equations.* Seizô Itô. Transl: Seizô Itô. Transl. of Math. Mono., V. 114. AMS, 1992, x + 225 pp, \$93. [ISBN 0-8218-4570-5]
- Partial Differential Equations, P.** *Systems of Evolution Equations with Periodic and Quasiperiodic Coefficients.* Yu. A. Mitropol'sky, A.M. Samoilenko, D.I. Martinyuk. Math. & Its Applic., V. 87. Kluwer Academic, 1993, xiv + 280 pp, \$126 pp. [ISBN 0-7923-2054-9]
- Partial Differential Equations, P.** *Nonlinear Potential Theory of Degenerate Elliptic Equations.* Juha Heinonen, Tero Kilpeläinen, Olli Martio. Oxford Math. Mono. Clarendon Pr, 1993, v + 363 pp, \$70. [ISBN 0-19-853669-0]
- Partial Differential Equations, P.** *Second International Conference on Mathematical and Numerical Aspects of Wave Propagation.* Eds: Ralph Kleinman, et al. SIAM, 1993, xi + 473 pp, \$69.50 (P). [ISBN 0-89871-318-8]

Proceedings of a 1993 conference at the University of Delaware.

Partial Differential Equations, P. *Floquet Theory for Partial Differential Equations*. Peter Kuchment. Oper. Theory: Adv. & Applic., V. 60. Birkhäuser, 1993, xiv + 350 pp, \$108.50. [ISBN 0-8176-2901-7]

Partial Differential Equations, P. *The Method of Newton's Polyhedron in the Theory of Partial Differential Equations*. S. Gindikin, L.R. Volevich. Math. & Its Applic., V. 86. Kluwer Academic, 1992, x + 266 pp, \$126. [ISBN 0-7923-2037-9]

Partial Differential Equations, P. *Algorithms for Elliptic Problems: Efficient Sequential and Parallel Solvers*. Marián Vajteršic. Math. & Its Applic., V. 58. Kluwer Academic, 1993, xviii + 292 pp, \$152. [ISBN 0-7923-1918-4]

Partial Differential Equations, P. *The Solution of a One-Dimensional Stefan Problem*. C. Vuik. CWI Tract, V. 90. Centrum voor Wiskunde en Informatica, 1993, 134 pp, Dfl. 40 (P). [ISBN 90-6196-419-9]

Partial Differential Equations, P. *Degenerate Elliptic Equations*. Serge Levendorskii. Math. & Its Applic., V. 258. Kluwer Academic, 1993, xi + 431 pp, \$186.50. [ISBN 0-7923-2305-X]

Partial Differential Equations, P. *Elliptic Boundary Problems for Dirac Operators*. Bernhelm Booß-Bavnbek, Krzysztof P. Wojciechowski. Math.: Theory & Applic. Birkhäuser, 1993, xviii + 307 pp, \$49.50. [ISBN 0-8176-3681-1]

Dynamical Systems, P. *Continuum Theory and Dynamical Systems*. Ed: Thelma West. Lect. Notes in Pure & Appl. Math., V. 149. Marcel Dekker, 1993, viii + 296 pp, \$115 (P). [ISBN 0-8247-9072-3] 19 papers, some expository, from a conference-workshop at the University of Southwestern Louisiana.

Dynamical Systems, P. *Analytic D-Modules and Applications*. Jan-Erik Björk. Math. & Its Applic., V. 247. Kluwer Academic, 1993, xiii + 581 pp, \$245. [ISBN 0-7923-2114-6]

Dynamical Systems, P. *Hamiltonian Mechanical Systems and Geometric Quantization*. Mircea Puta. Math. & Its Applic., V. 260. Kluwer Academic, 1993, viii + 278 pp, \$110. [ISBN 0-7923-2306-8]

Dynamical Systems, S(15-16), P, L. *The Dynamics of Ambiguity*. Giuseppe Caglioti. Springer-Verlag, 1992, xxi + 170 pp, \$59. [ISBN 0-387-52020-1] A slick, coffee-table-style book on symmetry and symmetry breaking in art and science. Translated from Italian into occasionally cumbersome English. BC

Numerical Analysis, P. *Knot Insertion and Deletion Algorithms for B-Spline Curves and Surfaces*. Eds: Ronald N. Goldman, Tom Lyche. SIAM, 1993, xiii + 197 pp, \$43.50 (P). [ISBN 0-89871-306-4] 7 papers providing modern perspectives on knot insertion and deletion algorithms for B-splines. AO

Numerical Analysis, P. *Acta Numerica 1993*. Cambridge Univ Pr, 1993, 326 pp, \$44.95. [ISBN 0-521-443563] 6 survey papers on path following, multivariate piecewise polynomials, numerical linear algebra, fluid dynamics, and multigrid methods.

Numerical Analysis, T*(14-15), L. *Elementary Numerical Computing with Mathematica*. Robert D. Skeel, Jerry B. Keiper. McGraw-Hill, 1993, xiv + 434 pp, \$60.16. [ISBN 0-07-057820-6] Interesting alternative to standard treatments. First two chapters treat error analysis and floating-point arithmetic. Remainder covers rootfinding, linear systems, interpolation, least-squares, numerical integration and differentiation, and ordinary differential equations. Appendices introduce Mathematica, survey basic mathematical topics. Authors contend that using Mathematica skirts programming problems, so reveals more mathematics. Worth a look. MPR

Numerical Analysis, P. *Guaranteed Accuracy in Numerical Linear Algebra*. S.K. Godunov, et al. Math. & Its Applic., V. 252. Kluwer Academic, 1993, xi + 535 pp, \$242. [ISBN 0-7923-2352-1]

Numerical Analysis, P. *Convergence of Iterations for Linear Equations*. Olavi Nevanlinna. Lect. in Math. Birkhäuser, 1993, vii + 177 pp, \$29 (P). [ISBN 0-8176-2865-7]

Operator Theory, P. *Generalized Vertex Algebras and Relative Vertex Operators*. Chongying Dong, James Lepowsky. Progress in Math., V. 112. Birkhäuser, 1993, ix + 202 pp, \$44.50. [ISBN 0-8176-3721-4]

Operator Theory, P. *Continuous and Discrete Fourier Transforms, Extension Problems and Wiener-Hopf Equations*. Ed: I. Gohberg. Oper. Theory, V. 58. Birkhäuser, 1992, viii + 214 pp, \$73. [ISBN 0-8176-2809-6] 7 papers, 4 based on talks from a September 1991 workshop at the University of Maryland.

Functional Analysis, P. *White Noise: An Infinite Dimensional Calculus*. Takeyuki Hida, et al. Math. & Its Applic., V. 253. Kluwer Academic, 1993, xiii + 516 pp, \$198. [ISBN 0-7923-2233-9]

Functional Analysis, P. *Positive Operators and Semigroups on Banach Lattices*. Eds: C.B. Huijsmans, W.A.J. Luxemburg. Kluwer Aca-

- demic, 1992, vii + 152 pp, \$92. [ISBN 0-7923-1964-8] Proceedings of the Caribbean Mathematics Foundation's 1990 conference. Reprinted from *Acta Applicandae Mathematicae*, Vol. 27, Nos. 1-2 (1992).
- Functional Analysis, P.** *Applied Theory of Functional Differential Equations.* V. Kolmanovskii, A. Myshkis. Math. & Its Applic., V. 85. Kluwer Academic, 1992, xv + 234 pp, \$99. [ISBN 0-7923-2013-1]
- Functional Analysis, P.** *Fourier Integrals in Classical Analysis.* Christopher D. Sogge. Tracts in Math., V. 105. Cambridge Univ Pr, 1993, x + 237 pp, \$39.95. [ISBN 0-521-43464-5]
- Functional Analysis, P.** *Functional Integrals: Approximate Evaluation and Applications.* A.D. Egorov, P.I. Sobolevsky, L.A. Yanovich. Math. & Its Applic., V. 249. Kluwer Academic, 1993, x + 418 pp, \$172. [ISBN 0-7923-2193-6]
- Functional Analysis, P.** *Invariant Function Spaces on Homogeneous Manifolds of Lie Groups and Applications.* M.L. Agranovskii. Transl. of Math. Mono., V. 126. AMS, 1993, x + 131 pp, \$71. [ISBN 0-8218-4604-3]
- Functional Analysis, P.** *Index Theory and Operator Algebras.* Eds: Jeffrey Fox, Peter Haskell. Contemp. Math., V. 148. AMS, 1993, vii + 190 pp, \$41 (P). [ISBN 0-8218-5152-7] Proceedings of the 1991 CBMS Regional Conference at the University of Colorado.
- Analysis, P.** *Lectures on Hermite and Laguerre Expansions.* Sundaram Thangavelu. Math. Notes, V. 42. Princeton Univ Pr, 1993, xv + 195 pp, \$22.50 (P). [ISBN 0-691-00048-4]
- Analysis, P.** *Difference Equations and Their Applications.* A.N. Sharkovsky, Yu. L. Maistrenko, E. Yu. Romanenko. Transl: D.V. Malyshev, P.V. Malyshev, Y.M. Pestryakov. Math. & Its Applic., V. 250. Kluwer Academic, 1993, xii + 358 pp, \$152. [ISBN 0-7923-2194-4] Modern theory, with many new results. Four parts: one-dimensional dynamical systems; difference equations with continuous time; differential-difference equations; boundary-value problems for hyperbolic systems of PDE's. AO
- Analysis, P.** *Commensurabilities among Lattices in $PU(1, n)$.* Pierre Deligne, G. Daniel Mostow. Annals of Math. Stud., No. 132. Princeton Univ Pr, 1993, 183 pp, \$19.95 (P); \$49.95. [ISBN 0-691-00096-4; 0-691-03385-4]
- Combinatorial Geometry, P.** *New Trends in Discrete and Computational Geometry.* Ed: János Pach. Algorithms & Combinatorics, V. 10. Springer-Verlag, 1993, xi + 339 pp, \$89. [ISBN 0-387-55713-X] 12 survey papers.
- Algebraic Geometry, P.** *Cyclic Homology.* Jean-Louis Loday. Grund. der math. Wissenschaften, B. 301. Springer-Verlag, 1992, xvii + 454 pp, \$149. [ISBN 0-387-53339-7]
- Algebraic Geometry, P. L.** *Mathematical Methods for CAD.* J.J. Risler. Cambridge Univ Pr, 1992, 196 pp, \$69.95; \$29.95 (P). [ISBN 0-521-43100-X; 0-521-43691-5] Introduction to B-splines for curves and surfaces and their applications. Uses tensor products, triangular Bezier patches; Bezier curves are treated as special cases of B-splines. Also treats algebraic problems, including properties of polynomials via formal computation. No exercises. MPR
- Algebraic Geometry, P.** *Mapping Class Groups and Moduli Spaces of Riemann Surfaces.* Eds: Carl-Friedrich Bödigheimer, Richard M. Hain. Contemp. Math., V. 150. AMS, 1993, xx + 372 pp, \$51 (P). [ISBN 0-8218-5167-5] Joint proceedings of workshops held in 1991 at the University of Göttingen and the University of Washington.
- Algebraic Geometry, P.** *Complex Analysis and Geometry.* Eds: Vincenzo Ancona, Alessandro Silva. Univ. Ser. in Math. Plenum Pr, 1993, xvi + 412 pp, \$85. [ISBN 0-306-44179-9] 16 papers reviewing recent research on the role of complex function theory in algebraic geometry, analytic geometry of manifolds and spaces, and differential geometry. Also includes a list of open problems.
- Algebraic Geometry, P.** *Algebraic Functions.* Kenkichi Iwasawa. Transl: Goro Kato. Transl. of Math. Mono., V. 118. AMS, 1993, xxii + 287 pp, \$131. [ISBN 0-8218-4595-0]
- Algebraic Geometry, P.** *Nilpotence and Periodicity in Stable Homotopy Theory.* Douglas C. Ravenel. Annals of Math. Stud., No. 128. Princeton Univ Pr, 1992, xiv + 209 pp, \$24.95 (P); \$69.50. [ISBN 0-691-02572-X; 0-691-08792-X]
- Algebraic Geometry, P.** *Introduction to Toric Varieties.* William Fulton. Annals of Math. Stud., No. 131. Princeton Univ Pr, 1993, xi + 157 pp, \$16.95 (P); \$32.50. [ISBN 0-691-00049-2; 0-691-03332-3]
- Algebraic Geometry, P.** *Tangents and Secants of Algebraic Varieties.* F.L. Zak. Transl. of Math. Mono., V. 127. AMS, 1993, vii + 164 pp, \$96. [ISBN 0-8218-4585-3]
- Differential Geometry, P.** *Einstein Metrics and Yang-Mills Connections.* Eds: Toshiki Mabuchi, Shigeru Mukai. Lect. Notes in Pure & Appl. Math., V. 145. Marcel Dekker, 1993, viii + 224 pp, \$99.75 (P). [ISBN 0-8247-9069-3]

Proceedings of the 27th Taniguchi International Symposium (Sanda, Japan, 1990).

Differential Geometry, P. *Constrained Mechanics and Lie Theory*. Robert Hermann. Interdisc. Math., V. 27. Math Sci Pr, 1992, 288 pp, \$95. [ISBN 0-915692-43-0] Nice introduction to modern differential geometric tools for constrained mechanical and variational systems. Basic tools are the theory of algebraic structure of vector fields defined by the Jacobi-Lie bracket and the structure of affine connections in subbundles of the tangent bundle, all developed using modern language and ideas. Also presents a modern treatment of Lie's exploitation of the symmetries of differential equations. Of interest to mathematicians working in mechanics, robotics, biomechanics. SP

Differential Geometry, P. *Minimal Surfaces*. Ed: A.T. Fomenko. Adv. in Soviet Math., V. 15. AMS, 1993, ix + 342 pp. [ISBN 0-8218-4116-5] 10 papers from the seminar on modern geometrical methods at Moscow State University.

Differential Geometry, P. *Harmonic Maps and Minimal Immersions with Symmetries: Methods of Ordinary Differential Equations Applied to Elliptic Variational Problems*. James Eells, Andrea Ratto. Annals of Math. Stud., V. 130. Princeton Univ Pr, 1993, 228 pp, \$19.95 (P); \$49.50. [ISBN 0-691-10249-X; 0-691-03321-8]

Differential Geometry, P. *Projective Differential Geometry of Submanifolds*. M.A. Akivis, V.V. Goldberg. Math. Lib., V. 49. North-Holland (US Distr: Elsevier Science), 1993, xi + 362 pp, \$128.50. [ISBN 0-444-89771-2]

Differential Geometry, P. *Differential Geometry*. Eds: Robert Greene, S.T. Yau. Proc. of Symposia in Pure Math., V. 54, Parts 1-3. AMS, 1993 [ISBN 0-8218-1493-1] set. *Partial Differential Equations on Manifolds*, Part 1, xxii + 560 pp, \$89 [ISBN 0-8218-1494-X]; *Geometry in Mathematical Physics and Related Topics*, Part 2, xxii + 655 pp, \$96 [ISBN 0-8218-1495-8]; *Riemannian Geometry*, Part 3, xxii + 710 pp, \$103. [ISBN 0-8218-1496-6] Proceedings of a 1990 AMS Summer Research Institute.

Differential Geometry, P. *Nonlinear Poisson Brackets: Geometry and Quantization*. M.V. Karasev, V.P. Maslov. Transl. of Math. Mono., V. 119. AMS, 1993, xi + 366 pp, \$170. [ISBN 0-8218-4596-9]

Algebraic Topology, P. *Algebraic Topology: Oaxtepec 1991*. Ed: Martin C. Tangora. Contemp. Math., V. 146. AMS, 1993, xviii + 481 pp, \$71 (P). [ISBN 0-8218-5162-4] Proceedings of the July 1991 conference.

Algebraic Topology, P. *Algebraic L-Theory and Topological Manifolds*. A.A. Ranicki. Tracts in Math., V. 102. Cambridge Univ Pr, 1992, 358 pp, \$69.95. [ISBN 0-521-42024-5] Algebraic L-theory is the algebraic K-theory of quadratic forms; it relates a manifold topology to its homotopy type (dimensions ≥ 5). SK

Algebraic Topology, P. *Differential Algebras in Topology*. David Anick. Res. Notes in Math., V. 3. AK Peters, 1993, xxv + 274 pp, \$49.50. [ISBN 1-56881-001-6]

Topology, P. *Embeddings and Immersions*. Masahisa Adachi. Transl: Kiki Hudson. Transl. of Math. Mono., V. 124. AMS, 1993, x + 183 pp, \$103. [ISBN 0-8218-4612-4]

Topology, P. *The Selected Works of J. Frank Adams, Volumes I-II*. Eds: J.P. May, C.B. Thomas. Cambridge Univ Pr, 1992, \$74.95 each. *Volume I*, xvi + 536 pp [ISBN 0-521-41063-0]; *Volume II*, xvi + 529 pp. [ISBN 0-521-41065-7]

Mathematical Modeling, P. *Fisheries: Control and Management via Application of Management Science (with a Comprehensive Annotated Bibliography of Applications)*. Eds: Bruce L. Golden, Edward A. Wasil. Amer. J. of Math. & Management Sci., V. 12, Nos. 2 & 3. American Sciences Pr, 1992, 150 pp, \$98.75 (P). [ISBN 0-935950-34-6] 5 papers on recent work.

Mathematical Modeling, P. *Forecasting the Health of Elderly Populations*. Eds: Kenneth G. Manton, Burton H. Singer, Richard M. Suzman. Ser. in Stat. Springer-Verlag, 1993, x + 371 pp, \$59. [ISBN 0-387-97953-0] 15 survey papers on methodological issues, forecasting techniques, effects of interventions on health care costs, and longitudinal research.

Control Theory, P. *Identification and Control in Systems Governed by Partial Differential Equations*. Eds: H.T. Banks, R.H. Fabiano, K. Ito. SIAM, 1993, ix + 234 pp, \$48.50 (P). [ISBN 0-89871-317-X] Proceedings of a 1992 AMS-IMS-SIAM Joint Summer Research Conference at Mt. Holyoke College.

Control Theory, P. *Lecture Notes in Control and Information Sciences-187: RoManSy 9*. Eds: A. Morecki, G. Bianchi, K. Jaworek. Springer-Verlag, 1993, xxxi + 438 pp, \$74 (P). [ISBN 0-387-19834-2] Proceedings of the 9th CISM-IFTOMM Symposium on Theory and Practice of Robots and Manipulators (Udine, Italy, 1992).

Stochastic Processes, P. *Schrödinger Equations and Diffusion Theory*. Masao Nagasawa. Mono. in Math., V. 86. Birkhäuser, 1993, 319 pp, \$99. [ISBN 0-8176-2875-4]

- Stochastic Processes, P.** *Some Aspects of Brownian Motion, Part I: Some Special Functionals.* Marc Yor. Lect. in Math. Birkhäuser, 1992, 136 pp, \$24.50 (P). [ISBN 0-8176-2807-X]
- Stochastic Processes, P.** *Sojourns and Extremes of Stochastic Processes.* Simeon M. Berman. Stat. & Prob. Ser. Wadsworth, 1992, xiv + 300 pp, \$59.95. [ISBN 0-534-13932-9]
- Stochastic Processes, P.** *Doebelin and Modern Probability.* Ed: Harry Cohn. Contemp. Math., V. 149. AMS, 1993, xii + 347 pp, \$49 (P). [ISBN 0-8218-5149-7] Proceedings of the 1991 conference "50 Years after Doebelin: Developments in the Theory of Markov Chains, Markov Processes, and Sums of Random Variables" held at Blaubeuren, Germany.
- Stochastic Processes, P.** *Gaussian Processes.* Takeyuki Hida, Masuyuki Hitsuda. Transl. of Math. Mono., V. 120. AMS, 1993, xv + 183 pp, \$99. [ISBN 0-8218-4568-3]
- Stochastic Processes, P.** *Contributions to Stochastics.* Ed: N. Venugopal. Wiley Eastern, 1992, xvi + 216 pp, Rs. 350. [ISBN 81-224-0417-0] 15 papers honoring K. Nagabhushanam.
- Statistical Methods, P.** *Drug Safety Assessment in Clinical Trials.* Ed: Gene Sogliero-Gilbert. Stat.: Textbooks & Mono., V. 138. Marcel Dekker, 1993, x + 437 pp, \$135. [ISBN 0-8247-8893-1] 15 papers on interpretation and analysis of safety data.
- Statistical Methods, P.** *Nonparametric Methods in Change-Point Problems.* B.E. Brodsky, B.S. Darkhovskiy. Math. & Its Applic., V. 243. Kluwer Academic, 1993, xi + 209 pp, \$90. [ISBN 0-7923-2122-7]
- Statistics, T(13: 1), C.** *Beginning Statistics: A to Z.* William Mendenhall. Duxbury Pr, 1993, xii + 525 pp, disk included [ISBN 0-534-19122-3]; *Flash Cards.* [ISBN 0-534-19122-3] Introductory text stresses concepts, statistical analyses, and their interpretation; downplays probability and hand calculations. Many examples and exercises stem from newspapers and research journals. Three appendices: data sets (available on disk), statistical tables, answers to selected exercises. Includes flash cards on statistical terminology and concepts. KB
- Computer Organization, P.** *The Cache Memory Book.* Jim Handy. Academic Pr, 1993, xviii + 269 pp. [ISBN 0-12-322985-5]
- Computer Systems, P.** *MS-DOS Command Summary for Versions 5 & 6.* SSC, 1993, 32 pp, (P). [ISBN 0-916151-64-6]
- Computer Systems, P.** *Motif Reference Manual, Volume Six B.* Paula M. Ferguson, David Brennan. O'Reilly & Assoc, 1993, xii + 908 pp, \$34.95 (P). [ISBN 1-56592-038-4]
- Computer Systems, P.** *Learning the Korn Shell.* Bill Rosenblatt. O'Reilly & Assoc, 1993, xxii + 338 pp, \$27.95 (P). [ISBN 1-56592-054-6]
- Computer Systems, P.** *PEXlib Programming Manual: 3D Programming in X.* Tom Gaskins. O'Reilly & Assoc, 1992, xlv + 1105 pp, \$44.95 (P). [ISBN 0-56592-028-7]
- Computer Systems, P.** *X Toolkit Intrinsics Programming Manual for Version 11, Volume Four.* Adrian Nye, Tim O'Reilly. O'Reilly & Assoc, 1992, xxxvi + 567 pp, \$34.95 (P). [ISBN 1-56592-003-1]
- Computer Systems, P.** *Software Portability with imake.* Paul DuBois. O'Reilly & Assoc, 1993, xxiii + 365 pp, \$27.95 (P). [ISBN 1-56592-055-4]
- Computer Systems, P.** *!%@:: A Directory of Electronic Mail Addressing and Networks.* Donnalyn Frey, Rick Adams. O'Reilly & Assoc, xvi + 443 pp, \$24.95 (P). [ISBN 1-56592-031-7]
- Computer Systems, P.** *X Window System User's Guide, Volume Three: OSF/Motif 1.2 Edition.* Valerie Quercia, Tim O'Reilly. O'Reilly & Assoc, 1993, xxxii + 924 pp, \$34.95 (P). [ISBN 1-56592-015-5]
- Computer Systems, P.** *Connecting to the Internet: A Buyer's Guide.* Susan Estrada. O'Reilly & Assoc, 1993, xv + 170 pp, \$15.95 (P). [ISBN 1-56592-061-9]
- Computer Systems, P.** *Learning the UNIX Operating System.* Grace Todino, John Strang, Jerry Peek. O'Reilly & Assoc, 1993, xv + 92 pp, \$9.95 (P). [ISBN 1-56592-060-0]
- Computer Systems, P.** *PEXlib Reference Manual: 3D Programming in X.* Ed: Steve Talbott. O'Reilly & Assoc, 1992, xxv + 551 pp, \$34.95 (P). [ISBN 1-56592-029-5]
- Computer Graphics, S*(16-17), P, L.** *Radiosity and Realistic Image Synthesis.* Michael F. Cohen, John R. Wallace. Academic Pr, 1993, xvi + 382 pp, \$49.95. [ISBN 0-12-178270-0] A complete presentation of tools for modern image synthesis. Radiosity improves the older technique of ray tracing. Ray tracing is viewer (and pixel) dependent, but radiosity is determined by an overall energy balance equation and so is viewer independent. In effect, radiosity takes into account indirect illumination of the scene. Requires substantial mathematics (e.g., differential geometry, numerical analysis,

measure theory), but much is developed in context. 54 color plates. MPR

Computer Graphics, S. C. *Modern Image Processing: Warping, Morphing, and Classical Techniques*. Christopher D. Watkins, Alberto Sadun, Stephen Marenka. Academic Pr, 1993, xix + 234 pp, \$49.95 (P), disk included. [ISBN 0-12-737860-X] Elementary introduction to filtering, image enhancement, and geometric transformations. Rather brief on mathematics behind these techniques, but surveys the subject. Includes a disk of code fragments (in C), and a short chapter on applications. MPR

Theory of Computation, P. *Formal Specification and Design*. L.M.G. Feijs, H.B.M. Jonkers. Tracts in Theoret. Comput. Sci., V. 35. Cambridge Univ Pr, 1992, xvi + 335 pp, \$44.95. [ISBN 0-521-43457-2]

Artificial Intelligence, P. *Computational Learning & Cognition: Proceedings of the Third NEC Research Symposium*. Ed: Eric B. Baum. SIAM, 1993, xi + 276 pp, \$42.50. [ISBN 0-89871-311-0]

Computer Science, P. *Computer Benchmarks*. Eds: Jack J. Dongarra, Wolfgang Gentzsch. Adv. in Parallel Comput., V. 8. North-Holland (US Distr: Elsevier Science), 1993, xiv + 349 pp, \$125.75. [ISBN 0-444-81518-X] 21 papers on performance prediction and measurement, compiler and database benchmarks, and techniques for benchmarking shared- and distributed-memory parallel computers.

Computer Science, P. *Virtual Reality: Applications and Explorations*. Ed: Alan Wexelblat. Academic Pr, 1993, xviii + 245 pp, \$39.95. [ISBN 0-12-745045-9] 10 essays on virtual reality in computer science, the arts, etc.

Computer Science, P. *Lectures on Parallel Computation*. Eds: Alan Gibbons, Paul Spirakis. Intern. Ser. on Parallel Comput., V. 4. Cambridge Univ Pr, 1993, 437 pp, \$49.95. [ISBN 0-521-41556-X] 15 papers on parallel computation, especially efficiency issues.

Computer Science, P. *Proceedings of the Sixth SIAM Conference on Parallel Processing for Scientific Computing, Volumes I-II*. Eds: Richard F. Sincovec, et al. SIAM, 1993, \$95 (P) [ISBN 0-89871-315-3]; *Volume I*, xix + 504 pp; *Volume II*, xix + 543 pp.

Computer Science, P. *Internet: Mailing Lists, 1993 Edition*. Eds: Edward T.L. Hardie, Vivian Neou. SRI Internet Inform. Ser. Prentice Hall, 1993, x + 356 pp, \$26 (P). [ISBN 0-13-327941-3]

Computer Science, P. *Proof Theory*. Eds: Peter Aczel, Stanley S. Wainer. Cambridge Univ Pr, 1992, x + 306 pp, \$49.95. [ISBN 0-

521-41413-X] 3 expository, 7 research papers from the 1990 Summer School and Conference on Proof Theory at Leeds University.

Applications (Biological Science), P. *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision*. Ed: Renaud Fortuner. Johns Hopkins Univ Pr, 1993, xiv + 560 pp, \$65. [ISBN 0-8018-4492-4] Proceedings of the ARTISYST workshop, an interdisciplinary conference on the application of modern computer methods to research in systematic biology, held at Napa, California, in 1990.

Applications (Fluid Dynamics), P. *Transonic Aerodynamics: Problems in Asymptotic Theory*. Ed: L. Pamela Cook. Frontiers in Appl. Math., V. 12. SIAM, 1993, x + 90 pp, \$26.50 (P). [ISBN 0-89871-310-2] 5 papers originally presented at a minisymposium at the second ICIAM meeting held in July 1991 in Washington, DC.

Applications (Fluid Dynamics), T(17-18: 1), P. *Numerical Modeling in Combustion*. Ed: T.J. Chung. Ser. in Comput. & Physical Processes in Mech. & Thermal Sci. Taylor & Francis, 1993, xxiv + 549 pp, \$195. [ISBN 0-89116-822-2] Numerical methods for reacting flow problems, especially combustion problems. Sections on laminar flows, turbulent flows, spray combustion. AO

Applications (Fluid Dynamics), S(16-18), P. *Bénard Cells and Taylor Vortices*. E.L. Koschmieder. Mono. on Mech. & Appl. Math. Cambridge Univ Pr, 1993, x + 337 pp, \$64.95. [ISBN 0-521-40204-2] Research on problems of Bénard convection, related areas of hydrodynamics. Graphs and photographs are almost as common as mathematical expressions. The bulk of the text is a very readable, informal, descriptive discussion. MU

Applications (Fluid Dynamics), P. *Theory of Laminar Film Condensation*. Tetsu Fujii. Springer-Verlag, 1991, xviii + 213 pp, \$69. [ISBN 0-387-97541-1]

Applications (Physics), P. *Quantum Field Theory*. Lowell S. Brown. Cambridge Univ Pr, 1992, xiv + 543 pp, \$100. [ISBN 0-521-400066]

Reviewers

KB: Karla Ballman, Macalester; JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; JD-B: John Dyer-Bennet, Carleton; SG: Steven Galovich, Carleton; DH: Deanna Haunsperger, St. Olaf; SK: Steve Kennedy, St. Olaf; AO: Arnold Ostebee, St. Olaf; SP: Samuel Patterson, Carleton; MPR: Matthew P. Richey, St. Olaf; MU: Milton Ulmer, Carleton; MW: Martha Wallace, St. Olaf.

Essential Mathematics from Cambridge

Numerical Mathematics —A Laboratory Approach

The late S. Breuer and G. Zwas
1993 280 pp. 44040-8 Hardcover \$49.95

The Higher Arithmetic

Sixth Edition

H. Davenport

1992 217 pp. 41998-0 Hardcover \$44.95
42227-2 Paper \$19.95

Primes and Programming

Computers and Number Theory

P. J. Giblin

1993 245 pp. 40182-8 Hardcover \$44.95
40988-8 Paper \$19.95

A Course of Pure Mathematics

Tenth Edition

G. H. Hardy

Cambridge Mathematical Library
1993 522 pp. 09227-2 Paper \$22.95

Elementary Theory of L-functions and Eisenstein Series

Haruzo Hida

*London Mathematical Society
Student Texts 26*

1993 398 pp. 43411-4 Hardcover \$69.95
43569-2 Paper \$29.95

Hyperbolic Geometry

Birger Iversen

*London Mathematical Society
Student Texts 25*

1993 312 pp. 43508-0 Hardcover \$54.95
43528-5 Paper \$22.95

Representations and Characters of Groups

Gordon James

and Martin Liebeck

1993 429 pp. 44024-6 Hardcover \$69.95
44590-6 Paper \$29.95

Exercises in Fourier Analysis

T. W. Korner

1993 395 pp. 43276-6 Hardcover \$54.95
43849-7 Paper \$22.95

Computational Geometry in C

Joseph O'Rourke

1994 320 pp. 44034-3 Hardcover \$59.95
44592-2 Paper \$24.95

Hilbert Space

Compact Operators and
the Trace Theorem

J. R. Retherford

*London Mathematical Society
Student Texts 27*

1993 143 pp. 41884-4 Hardcover \$44.95
42933-1 Paper \$19.95

Elementary Probability

David Stirzaker

1994 500 pp. 42028-8 Hardcover \$64.95
42183-7 Paper \$24.95

How To Prove It

A Structured Approach

Daniel J. Velleman

1994 304 pp. 44116-1 Hardcover \$49.95
44663-5 Paper \$19.95

Available in bookstores or from

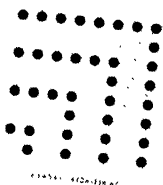
CAMBRIDGE
UNIVERSITY PRESS

40 West 20th St., N.Y., NY 10011-4211
Call toll-free 800-872-7423
MasterCard/VISA accepted.
Prices subject to change.

Proofs Without Words

Exercises in Visual Thinking

Roger B. Nelsen



LEARNING MATERIALS / NUMBER 1
THE MATHEMATICAL ASSOCIATION OF AMERICA

Just what are “proofs without words?” First of all, most mathematicians would agree that they certainly are not “proofs” in the formal sense. Indeed, the question does not have a simple answer. Proofs without words are generally pictures or diagrams that help the reader see *why* a particular mathematical statement may be true, and *how* one could begin to go about proving it. While in some proofs without words an equation or two may appear to help guide that process, the emphasis is clearly on providing *visual* clues to stimulate mathematical thought. Proofs without words bear witness to the observation that often in the English language to *see* means to *understand*, as in “to see the point of an argument.”

Proofs without words have a long history. In this collection you will find modern renditions of proofs from ancient China, classical Greece, twelfth-century India—even one based on a published proof by a former President of the United States! However, most of the proofs are more recent creations, and many are taken from the pages of MAA journals.

The proofs in this collection are arranged by topic into six chapters: Geometry and Algebra; Trigo-

nometry, Calculus and Analytic Geometry; Inequalities; Integer Sums; Sequences and Series; and Miscellaneous. Teachers will find that many of the proofs in this collection are well suited for classroom discussion and for helping students to think visually in mathematics.

The readers of this collection will find enjoyment in discovering or rediscovering some elegant visual demonstrations of certain mathematical ideas that teachers will want to share with their students. Readers may even be encouraged to create new “proofs without words.”

160 pp., Paperbound, 1993

ISBN 0-88385-700-6

List: \$27.50 MAA Member: \$22.00

Catalog Number PWW

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty.	Catalog Number	Price
------	----------------	-------

_____	_____	_____
_____	_____	_____

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

Was This The Last Time You Bought Insurance?



Face it—it's been a long time. A lot has changed since then. Your family. Maybe your job. And more than likely, the amount and types of coverage you need from your insurance program. That's why you need insurance that can easily adapt to the way your life changes—MAA Group Insurance Program.

We Understand You.

Finding an insurance program that's right for you isn't easy. But as a member of MAA, you don't have to go through the difficult and time consuming task of looking for the right plans—we've done that work for you. What's more, you can be sure the program is constantly being evaluated to better meet the needs of our members.

We're Flexible.

Updating your insurance doesn't have to be a hassle. With our plans, as your needs

change, so can your coverage. Insurance through your association is designed to grow with you—
it even moves with you
when you change jobs.

We're Affordable.

What good would all these benefits be if no one could afford them? That's why we offer members the additional benefit of reasonable rates, negotiated using our group purchasing power. Call 1 800 424-9883 (in Washington, D.C., (202) 457-6820) between 8:30 a.m. and 5:30 p.m. Eastern Time for more information about these insurance plans offered through MAA:

Term Life • Disability Income Protection •
Comprehensive HealthCare • Excess Major
Medical • In Hospital • High-Limit Accident

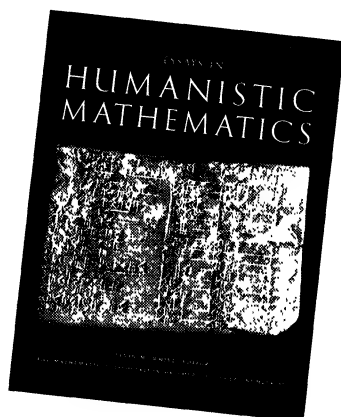
MAA Insurance

Designed for the way you live today.
And tomorrow.

This plan is administered by Seabury & Smith.

Essays in Humanistic Mathematics

Alvin White, Editor



A dazzling array of essayists reveal humanistic mathematics in this volume, and in so doing go beyond the facts, formulas, and algorithms that most students associate with mathematics to a presentation of mathematics as an intellectual discipline with a human perspective and a significant history. Humanistic mathematics challenges dogmatic teaching styles that expect students to parrot the lecturer. It demands creativity from both the teacher and student.

Teaching mathematics humanistically seeks to place the student more centrally in the position of inquirer than is generally the case, while at the same time acknowledging the emotional climate of the activity of learning mathematics. This type of teaching encourages students to learn from each other and to better understand mathematics as socially constructed knowledge, rather than as an arbitrary discipline.

Teaching humanistic mathematics brings the focus less upon the nature of the teaching and learning environment and more upon the need to reconstruct the curriculum and the discipline of mathematics itself. This reconstruction relates mathematical discoveries to personal courage, discovery to verification, mathematics to science, truth to utility, and

mathematics to the culture in which it is embedded.

The humanistic mathematics movement, which began as the personal vision of a few, has now become a major part of mathematical culture. What was viewed with skepticism is now accepted and expected. Humanistic mathematics is not a new discovery. It is a recent rediscovery of ideas that go back to Plato. It has provided a vocabulary for previously unarticulated concepts and approaches.

The essays in this volume illustrate and help to define humanistic mathematics. The variety and scope indicate the richness and fruitfulness of the concept. Although each essay is independent, a sense of unity emerges. A glimpse at the table of contents will give you an idea of the excitement and range of the ideas presented.

212 pp., Paperbound, 1993

ISBN 0-88385-089-3

List: \$24.00

Catalog Number NTE-32

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty.	Catalog Number	Price
------	----------------	-------

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

Graphing Calculator Workshops and Short Courses for College Faculty
Organized by Frank Demana and Bert Waits of The Ohio State University

We are seeking host sites for graphing calculator workshops or short courses to be given in the spring, summer, and fall of 1994 and 1995. We have a network of dynamic college instructors available to present one, one and one half, or two-day workshops or three-day to one-week "short courses" on using Texas Instruments graphing calculators to enhance the teaching and learning of college mathematics. Topics may be selected from developmental mathematics through calculus and differential equations. Participants will receive extensive "hands-on" instruction from experienced colleagues and will receive many useful self study materials. We have five summer 1994 short courses scheduled.

Short Courses Offered Summer 1994

June 1 - 5, 1994 **Rochester, NY**
Host Site: *Rochester Institute of Technology*
June 6 - 10, 1994 **Costa Mesa, CA**
Host Site: *Orange Coast College*
July 5 - 9, 1994 **Edison, NJ**
Host Site: *Middlesex County College*
July 11 - 15, 1994 **Columbus, OH**
Host Site: *Columbus State Community College*
July 13 - 17, 1994 **Orlando, FL**
Host Site: *Valencia Community College*
\$100 registration fee for the one-week short course.

To obtain more information, write:
Bert Waits and Frank Demana
College Graphing Calculator Workshop/Short
Course Program
Department of Mathematics
The Ohio State University
231 West 18th Avenue
Columbus, OH 43210

**For College
Mathematics
Teachers**

**A SOURCE BOOK FOR
COLLEGE MATHEMATICS
TEACHING**

Alan Schoenfeld, Editor.
Prepared by the Committee on the
Undergraduate Teaching of Mathematics

Do you want a broader, deeper, more successful mathematics program? This Source Book points to the resources and perspectives you need.

This book provides the means for improving instruction, and describes the broad spectrum of mathematical skills and perspectives our student should develop. The curriculum recommendations section shows where to look for reports and course re-

sources that will help you in your teaching. Extensive descriptions of advising programs that work is included, along with suggestions for teaching that describe a wide range of instructional techniques. You will learn about how to use computers in your teaching, and how to evaluate your performance as well as that of your students.

Every faculty member concerned about teaching should read this book. Every administrator with responsibility for the quality of mathematics programs should have a copy.

80 pp., 1990, Paper,
ISBN 0-88385-068-0

List \$10.00

Catalog Number SRCE

ORDER FROM



**The Mathematical Association
of America**

1529 Eighteenth Street, N.W.
Washington, D.C. 20036

The Solutions to Your Problems.



The History of Mathematics

Volume III

Eberhard Knobloch and David E. Rowe

This volume contains nine essays dealing with historical issues of mathematics. The topics covered span three different approaches to the history of mathematics. The first section addresses the historiographical and philosophical issues involved in determining the meaning of mathematical history. The second section traces the convoluted development of the ideas of differential geometry and analysis. The third section discusses the structure and interaction of mathematical communities through studies of the social fabric of the mathematical communities of the U.S. and China.

April 1994, c. 453 pp., \$55.00 (tentative)/ISBN: 0-12-599663-2



Topics in Geometry

Robert Bix

Topics in Geometry opens with a brief review of elementary geometry before proceeding to advanced material. Topics covered include advanced Euclidean and non-Euclidean geometry, division ratios and triangles, transformation geom-

etry, projective geometry, conic sections, and hyperbolic and absolute geometry. The text includes over 800 illustrations and extensively worked-out exercises.

October 1993, 538 pp., \$59.95/ISBN: 0-12-102740-6

Fifth Edition!

Table of Integrals, Series, and Products

FIFTH EDITION

I.S. Gradshteyn and I.M. Ryzhik

Alan Jeffrey, Editor

October 1993, 1204 pp., \$54.95/ISBN: 0-12-294755-X

Computer Algebra

Systems and Algorithms for Algebraic Computation

SECOND EDITION

J.H. Davenport, Y. Siret, and E. Tournier

This updated Second Edition provides a comprehensive review of the subject and contains excellent references to fundamental papers and worked examples. In addition, the book includes an appendix describing the use of one particular algebra system—REDUCE.

1993, 298 pp., \$45.00/ISBN: 0-12-204232-8

Order from your local bookseller or directly from

ACADEMIC PRESS, Inc.

A Division of Harcourt Brace & Company

Order Fulfillment Dept. DM17915

6277 Sea Harbor Drive, Orlando, FL 32887

CALL TOLL FREE

1-800-321-5068

FAX **1-800-336-7377**

Prices subject to change without notice.

© 1994 by Academic Press, Inc. All Rights Reserved

Under New Editorial Direction

Historia Mathematica

Editor

David E. Rowe

Universität Mainz, Germany

Managing Editor

Karen H. Parshall

University of Virginia, Charlottesville

Historia Mathematica is concerned with the history of all aspects of the mathematical sciences in all parts of the world and all historical periods. The journal publishes occasional biographies of mathematicians and historians, studies of organizations and institutions, essays on historiography, and articles on the interactions among all facets of mathematical activity and other aspects of culture and society.

Published by Academic Press, Inc., a Division of Harcourt Brace & Company, for the Division of the History of Science of the International Union of the History and Philosophy of Science

Volume 21 (1994), 4 issues

ISSN 0315-0860

In the U.S.A. and Canada: \$127.00

All other countries: \$156.00

Free Sample Copies and privileged personal rates are available on request. For more information, please contact:

ACADEMIC PRESS, Inc.

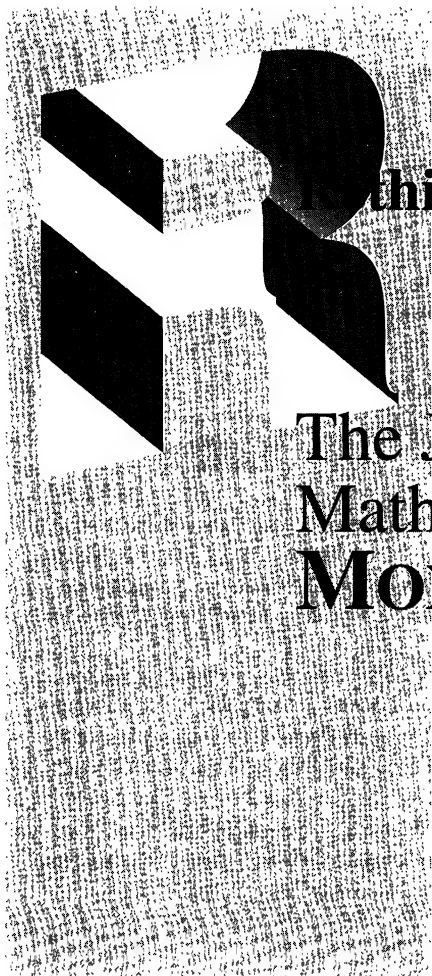
Journal Promotion Department

525 B Street, Suite 1900

San Diego, CA 92101-4495, U.S.A.

1-800-894-3434

All prices are in U.S. dollars and are subject to change without notice. Canadian customers: Please add 7% Goods and Services Tax to your order.



Think Your Approach to Teaching Mathematics With the Help of The Journal for Research in Mathematics Education **MONOGRAPH SERIES**

Learning and Mathematics Games, by G. W. Bright, J. G. Harvey, and M. M. Wheeler. Describes elementary and secondary classroom studies that investigate the use of games to promote mathematics learning. Softcover, 1985, 199 pp., ISBN 0-87353-233-3, #357, \$9.00*.

Learning to Add and Subtract, by T. A. Romberg and K. F. Collis. Features studies in grades 1–3 of how differences in cognitive capacity affect the way children learn to add and subtract. Softcover, 1987, 177 pp., ISBN 0-87353-242-2, #372, \$9.00*.

The van Hiele Model of Thinking in Geometry among Adolescents, by D. Fuys, D. Geddes, and R. Tischler. Presents analyses of interviews with sixth and ninth graders to investigate how they learn geometry. Softcover, 1988, 208 pp., ISBN 0-87353-266-X, #394, \$8.00*.

Constructivist Views on the Teaching and Learning of Mathematics, by R. B. Davis, C. A. Mahler, and N. Noddings. Considers the background of constructivism and explores the process of mathematical thinking. Describes how teachers can promote mathematical activity in children. Softcover, 1990, 210 pp., ISBN 0-87353-300-3, #479, \$10.50*.

An Ethnographic Study of the Mathematical Ideas of a Group of Carpenters, by W. L. Millroy. Documents the varied mathematical concepts used by carpenters everyday. Shows how the concepts of congruence, symmetry, proportion, and spatial visualization are routinely applied. Softcover, 1992, 210 pp., ISBN 0-87353-341-0, #535, \$7.50*.

Rethinking Elementary School Mathematics: Insights and Issues, by T. Wood, P. Cobb, E. Yackel, and D. Dillon. Chronicles a psychological and sociological investigation of children's mathematical learning in a second-grade classroom over the course of one year. Describes problem-solving episodes with children in pairs, as well as teacher-directed whole-class discussions. Softcover, 1993, 122 pp., ISBN 0-87353-362-3, #547, \$7.00*.



**NATIONAL COUNCIL OF
TEACHERS OF MATHEMATICS**

1906 Association Drive
Reston, VA 22091-1593
(703) 620-9840, fax (703) 476-2970

**For publication orders and
memberships only, (800) 235-7566**

NEW!

Knot Theory

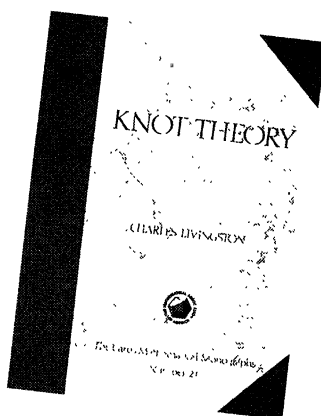
Charles Livingston

I learned more about knots after an hour with the book than I thought I could, and I am glad that it is here on my desk so that I may spend more time with it and, I hope, learn more.
—Paul Halmos

Knot Theory, a lively exposition of the mathematics of knotting, will appeal to a diverse audience from the undergraduate seeking experience outside the traditional range of studies to mathematicians wanting a leisurely introduction to the subject. Graduate students beginning a program of advanced study will find a worthwhile overview, and the reader will need no training beyond linear algebra to understand the mathematics presented.

Over the last century, knot theory has progressed from a study based largely on intuition and conjecture into one of the most active areas of mathematical investigation. **Knot Theory** illustrates the foundations of knotting as well as the remarkable breadth of techniques it employs—combinatorial, algebraic, and geometric.

The interplay between topology and algebra, known as algebraic topology, arises early in the book, when tools from linear algebra and from basic group theory are introduced to study the properties of knots, including the unknotting number, the braid index, and the bridge number. Livingston guides you through a general survey of the topic showing how to use the techniques of linear algebra to address some sophisticated problems, including one of mathematics' most beautiful topics, symmetry. The book closes with a discussion of high-dimensional knot theory and a presentation of some



of the recent advances in the subject—the Conway, Jones and Kauffman polynomials. A supplementary section presents the fundamental group, which is a centerpiece of algebraic topology.

An extensive collection of exercises is included. Some problems focus on details of the subject matter; others introduce new examples and topics illustrating both the wide range of knot theory and the present borders of our understanding of knotting. All are designed to offer the reader the experience and pleasure of working in this fascinating area.

264 pp., Hardbound, 1993

ISBN 0-88385-027-3

List: \$31.50 MAA Member: \$25.00

Catalog Number CAM-24

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036
1-(800) 331-1622 (202)-387-5200



Membership Code

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty.	Catalog Number	Price
------	----------------	-------

_____	_____	_____
-------	-------	-------

_____	_____	Total \$ _____
-------	-------	----------------

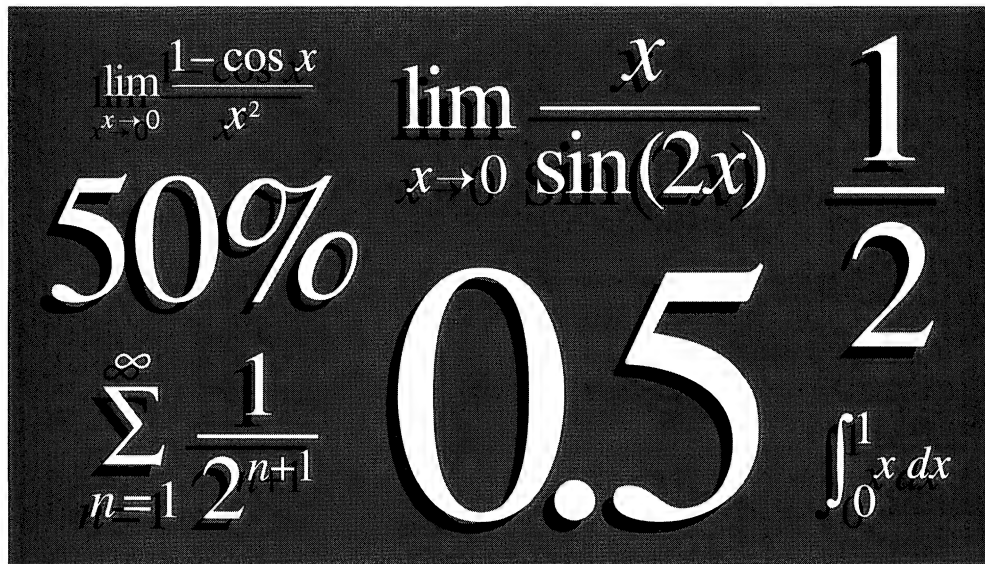
Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

**No matter how you
express it, it still means
DERIVE® is half price.**



DERIVE ➡

The *DERIVE A Mathematical Assistant* program lets you express yourself symbolically, numerically and graphically, from algebra through calculus, with vectors and matrices too—all displayed with accepted math notation, or 2D and 3D plotting. *DERIVE* is also easy to use and easy to read, thanks to a friendly, menu-driven interface and split or

overlay windows that can display both algebra and plotting simultaneously. Better still, *DERIVE* has been praised for the accuracy and exactness of its solutions. But, best of all the suggested retail price is now only \$125. Which means *DERIVE* is now half price, no matter how you express it.

System requirements

DERIVE: MS-DOS 2.1 or later, 512K RAM, and one 3½" disk drive. Suggested retail price now **\$125 (Half off!)**.

DERIVE ROM card: Hewlett Packard 95LX & 100LX Palmtop, or other PC compatible ROM card computer. Suggested retail price now **\$125!**

DERIVE XM (eXtended Memory): 386 or 486 PC compatible with at least 2MB of *extended* memory. Suggested list price now \$250!

DERIVE is a
registered trademark of
Soft Warehouse, Inc.

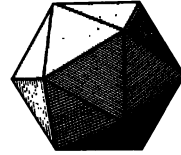


Soft Warehouse
HONOLULU • HAWAII

Soft Warehouse, Inc. • 3660 Waiialae Ave.
Ste. 304 • Honolulu, HI, USA 96816-3236
Ph: (808) 734-5801 • Fax: (808) 735-1105

**The American
Mathematical Monthly**

Volume 101 Number 3 / MARCH 1994
(ISSN 0002-9890)



Contents

ARTICLES

- A Marvelous Proof / FERNANDO Q. GOUVÊA 203
Triangulating the Circle, at Random / DAVID ALDOUS 223
Hypatia and Her Mathematics / MICHAEL A. B. DEAKIN 234
Calculus II and Euler Also (with a Nod to Series Integral
Remainder Bounds) / RICHARD BARSHINGER 244
A Focusing Property of the Ellipse / MARC FRANTZ 250
-

FEATURES

COMMENTS 202

NOTES

- Reflections Can Be Trapped / ROBERTO PEIRONE 259
Euler's Theorem / KATHERINE HEINRICH
and PETER HORAK 260

PICTURE PUZZLE 261

THE COMPUTER SCIENCE SAMPLER

- Universal Traversal Sequences / JOAN FEIGENBAUM
and NICK REINGOLD 262

THE EVOLUTION OF ...

- What Are Algebraic Integers and What Are They For?/
JOHN STILLWELL 266

THE AUTHORS 271

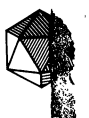
PROBLEMS AND SOLUTIONS 273

REVIEWS

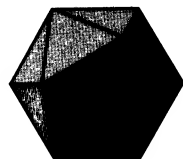
- Revolutions in Mathematics*. Edited by Donald Gillies /
MICHAEL MAHONEY 283

TELEGRAPHIC REVIEWS 288

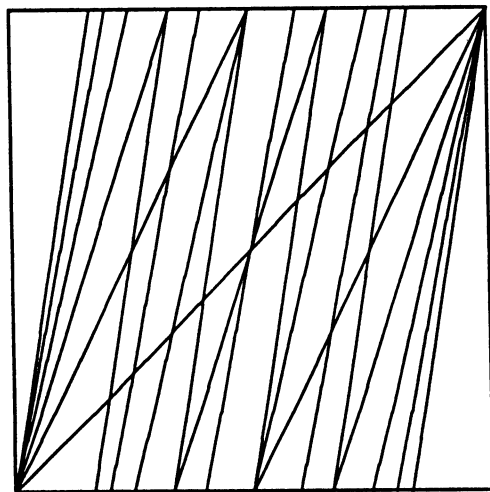
THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
WASHINGTON, D.C. 20036



The American Mathematical Monthly



Volume 101, Number 4 / APRIL 1994



NOTICE TO AUTHORS

The *Monthly* publishes articles, notes, and other features about mathematics and the profession. The readership of the *Monthly* is intended to include everybody who is mathematically inclined, including of course professional mathematicians and students of mathematics at all collegiate levels. While no single article or feature is likely to appeal to everyone, material should interest and be accessible to a large number of readers. This is the most important criterion for acceptance.

Articles may be expositions of old results or presentations of new ones. They may concern all of mathematics or one small area, a broad development or a single application, historical reminiscences or one important event. While some articles may contain the author's new research, the novelty of material and generality of the results is far less important than the clarity of exposition and general interest. Discussing one illuminating case of a well known result is far better than providing all the details of an obscure but new proposition. Articles in the *Monthly* are supposed to inform and to entertain; they are meant to be read rather than archived.

Notes are short and possibly informal articles. A note may concern a clever new proof of an old theorem, a novel way to present tired material, or a lively discussion of a philosophical (but still mathematical) issue. Also, any topic is suitable, so long as it is related to mathematics. Because a note is short, the first few sentences are the most important part: They should explain the purpose and invite the reader in. Photographs or diagrams often will attract the reader's attention.

All articles and notes should be sent to the editor:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Please send 3 copies, typewritten on only one side of the paper. Illustrations should be carefully drawn on separate sheets of paper in black ink; the original should be without lettering and two copies should have appropriate captions and lettering indicated.

Proposed problems or solutions should be sent to:

RICHARD BUMBY,
P.O. Box 10971
New Brunswick, NJ 08906-0971.

Please send 2 copies of all material, typewritten if possible.

Letters to the Editor, both for publication and for private reading, should be sent to the Editor at the address given above. Comments, including criticisms, are welcome, as are all suggestions for making the *Monthly* a lively, entertaining, and informative journal.

EDITOR:

JOHN H. EWING

ASSOCIATE EDITORS:

PETER BORWEIN	FRED KOCHMAN
RICHARD BUMBY	CATHERINE MCGEOCH
DENNIS DETURCK	RICHARD NOWAKOWSKI
UNDERWOOD DUDLEY	ARNOLD OSTELEE
JOHN DUNCAN	LEE RUBEL
JOAN FERRINI-MUNDY	ABE SHENITZER
JOSEPH GALLIAN	LYNN STEEN
STEVEN GALOVICH	STAN WAGON
RICHARD GUY	DOUGLAS WEST
DARRELL HAILE	HERBERT WILF
PAUL HALMOS	SANDY ZABELL
JOAN HUTCHINSON	PAUL ZORN

EDITORIAL ASSISTANT:

MISTY CUMMINGS

STAFF ARTIST:

MIKE CAGLE

Reprint permission:

MARCIA P. SWARD, Executive Director

Advertising Correspondence:

Ms. ELAINE PEDREIRA, Advertising Manager

Subscription correspondence, change of address, and other inquiries:

Membership / Subscriptions Department

All at the address:

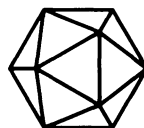
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT. Copyrighted by the Mathematical Association of America (Incorporated), 1994, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source. Second class postage paid at Washington, DC, and additional mailing offices. **Postmaster:** Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.

**The American
Mathematical Monthly**

Volume 101 Number 4 / APRIL 1994
(ISSN 0002-9890)



Contents

ARTICLES

**Pizza Slicing, Phi's and the Riemann Hypothesis / EDWARD A. BENDER,
OREN PATASHNIK, and HOWARD RUMSEY, JR. 307**

**Rational Periodic Points of the Quadratic Function $Q_c(x) = x^2 + c$ /
RALPH WALDE and PAULA RUSSO 318**

**Fréchet vs. Carathéodory / ERNESTO ACOSTA G.
and CESAR DELGADO G. 332**

Odd Magic Powers / A. C. THOMPSON 339

**Mathematicians, Including Undergraduates, Look at Soap Bubbles /
FRANK MORGAN 343**

FEATURES

COMMENTS 306

NOTES

**A Proof of Dilworth's Chain Decomposition Theorem /
FRED GALVIN 352**

**On Intervals, Transitivity = Chaos / MICHEL VELLEKOOP
and RAOUL BERGLUND 353**

**Proof of a Mixed Arithmetic-Mean, Geometric-Mean Inequality /
KIRAN KEDLAYA 355**

UNSOLVED PROBLEMS

ApSimon's Mints Problem 358

THE AUTHORS 360

PROBLEMS AND SOLUTIONS 362

REVIEWS

***Mathematics of the 19th Century: Mathematical Logic, Algebra,
Number Theory, Probability Theory*, edited by A. N. Kolmogorov
and A. P. Yushkevich / KAREN HUNGER PARSHALL 369**

***Calculus Gems: Brief Lives and Memorable Mathematics.*
By George F. Simmons / DAVID J. PENGELLEY 374**

TELEGRAPHIC REVIEWS 381

COMMENTS

"I have been invited to discuss further the subject which you had under consideration last year, namely, the problem of turning Doctors of Philosophy into efficient teachers. There are some of you, I see by the printed reports of your earlier meeting, who regard the ability to teach as a natural gift. There are apparently many of you who are persuaded that courses in pedagogy cannot contribute materially to the improvement of a graduate of a university. In the face of such settled views, backed up by the success that many of you have obtained in the teaching profession, it seems to be a bold and from some point of view a useless undertaking to come before you as I must with the assertion that ability to teach is not a natural gift and that every Doctor of Philosophy would be improved by a careful consideration in a scientific and historical way of the problems of education

"Until very recently university and college organizations have been based on the opinion expressed by some of you that teaching cannot be made a subject of special study and instruction. Gradually, however, a change is being worked out before our eyes. In spite of opposition and indifference, courses in education are being organized even in the most conservative institutions

"The first fact which I wish to point out is that there is a great deal of poor teaching within our universities and colleges, and this fact can be traced to a neglect upon the part of academic men and women of the form in which they arrange their material

"It is widely recognized that such a neglect of form as I have indicated appears in much of our university lecturing. The assumption of the ordinary university lecturer is that if he presents a certain body of material so organized that it seems to him to be fairly coherent and logical in its character, it is a matter of small moment whether it appeals to the students because of its literary form or whether it is easily intelligible to them because of its careful adjustment to their present stage of development

"The elective system has in part overcome this attitude and there is a very much more general conviction on the part of college instructors now than there was a generation ago that the demands of the student that the material shall be presented in clear and coherent form should be met. An instructor who is in competition with the other members of his own department for students in his course is likely to recognize the importance of preparing his material as clearly as possible for presentation to his class

"As matters now stand, whatever the ideal held by the university world in regard to the high calling of the doctor in the field of pure research, the fact is that the great majority of the doctors now turned out in this country are nominated and pushed by their respective departments, for teaching positions, which they must fill successfully or be counted as failures by all who measure the ratio of results accomplished to tasks undertaken Success in teaching is well-nigh indispensable."

Professor Charles Hubbard Judd

Director, University of Chicago Education School
as reported in The November MONTHLY, 16(1909)

Pizza Slicing, Phi's, and the Riemann Hypothesis

Edward A. Bender, Oren Patashnik, and Howard Rumsey, Jr.

1. INTRODUCTION. Place a unit-square pizza so that its lower-left corner is at the origin and its upper-right corner is at the point $(1, 1)$. Run a pizza knife along all lines having an integer y -intercept and a positive-integer slope that doesn't exceed N . How many slices $S(N)$ are formed, and how many sides might a slice have?

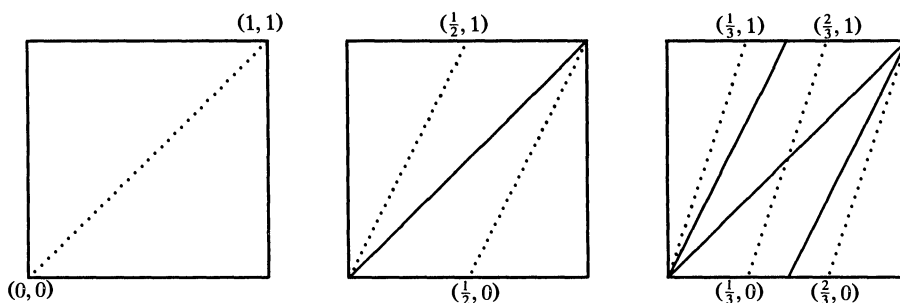


Figure 1. The 2, 4, and 8 slices formed when $N = 1, 2$, and 3 .

FIGURE 1 shows that when $N = 1$ the only relevant line is $y = x$, and two triangular slices are formed. When $N = 2$ there are two additional lines, $y = 2x$ and $y = 2x - 1$, and there are now four triangular slices (henceforth called simply "triangles"). When $N = 3$ there are eight slices—six triangles and two quadrilaterals. FIGURE 2 shows the 22 triangles and 14 quadrilaterals for $N = 6$.

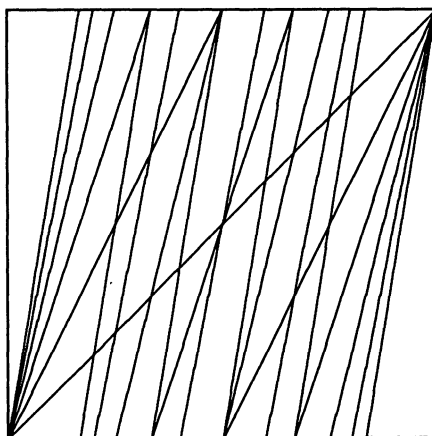


Figure 2. The 36 slices (22 triangles, 14 quadrilaterals) when $N = 6$.

David desJardins, who posed the general problem, also asked in particular whether the slices are always triangles or quadrilaterals. This paper answers “yes” to that question, giving formulas for the number of each that involve a double sum of Euler’s φ -function, and it shows that asymptotically there are N^3/π^2 slices in all (half triangles, half quadrilaterals), with an error term that involves the Riemann hypothesis. Along the way, the paper gives a fairly complete history of the somewhat controversial asymptotics of $\sum_{1 \leq k \leq x} \varphi(k)$.

2. ESTIMATING SLICES. First some terminology. From now on a *slice* refers to a region of the unit square, rather than the corresponding part of the pizza. A *segment* is the part of one of our lines lying in the unit square; we’ll talk about a segment as if it were of the form $y = Ax - B$, even though technically it is the corresponding line that’s of that form. And \mathcal{S} is the set of segments having nonzero length. So $S(N)$ is just the number of slices into which the set \mathcal{S} partitions the unit square.

To get a crude upper bound on $S(N)$ we start with an easy question: How many segments are in \mathcal{S} ? One of slope 1, two of slope 2, and so on, hence $\sum_{k=1}^N k = N(N+1)/2 = O(N^2)$ segments total. Now it’s well known that l straight lines partition the plane into at most $1 + l(l+1)/2 = O(l^2)$ regions. (There are exactly that many regions, it turns out, if the lines are in general position; that is, if no two lines coincide or are parallel, and if no three lines intersect at a point.) To deduce the $O(l^2)$ bound, notice that when we add a k th (distinct) line to a collection of $k-1$ lines in the plane, the number of additional regions formed is one more than the number of distinct intersection points the new line makes with the old lines; since there are at most $k-1$ such points, there are at most k additional regions, hence after we’ve added l lines there are at most $1 + \sum_{k=1}^l k$ regions, as claimed. (Jacob Steiner [14] proved this formula in the early 19th century; Graham et al. [4, Chapter 1] discuss it, along with some generalizations.) Therefore, since the number of l -line regions in the plane is an upper bound for the number of l -segment slices in the unit square, the $O(N^2)$ segments of \mathcal{S} yield $O(N^4)$ slices.

But $O(N^4)$ is not a very good bound, because the segments of \mathcal{S} aren’t even close to being in general position—some segments (all k of slope k) are parallel, and many can have a mutual intersection point (there are $\lfloor N/2 \rfloor$ segments passing through the point $(\frac{1}{2}, \frac{1}{2})$ for example).

We can improve the bound on $S(N)$ by a factor of N , using one more observation: A segment of slope k intersects at most one of the segments of slope j , for each $j < k$, because the projection onto the x axis of the segment of slope k is properly contained in the projection of a segment of slope j that it intersects, and because the projections of the j segments of slope j are pairwise disjoint. So when we start with the empty square and add the segment of slope 1, then the segments of slope 2, then slope 3, and so on, a segment of slope k intersects at most $k-1$ other segments and thus increases the number of slices by at most k . Hence collectively the segments of slope k add at most k^2 slices, giving at most $\sum_{k=1}^N k^2 = N^3/3 + O(N^2)$ slices total. (As noted in the introduction, $S(N)$ ’s actual leading term has π^2 in place of 3 in the denominator.)

3. COUNTING SLICES. We now derive an exact expression for $S(N)$. Let $\varphi(k)$ be Euler’s totient function—the number of integers in the interval $[0..k)$ that are prime to k —and let Φ be the sum of φ :

$$\Phi(x) = \sum_{1 \leq k \leq x} \varphi(k). \quad (1)$$

The exact expression for $S(N)$ involves the sum of Φ :

$$S(N) = 1 + \sum_{1 \leq k \leq N} \Phi(k), \quad \text{for } N \geq 0. \quad (2)$$

To derive this expression, we start as before with the empty square, which comprises a single slice, and we add the segments in order of increasing slope. We'll see that the k segments of slope k collectively increase the slice count by $\Phi(k)$, from which Equation (2) follows.

The lines-in-the-plane argument above shows that when we add a single segment to the unit square, the slice count increases by one more than the number of points of intersection, excluding points on the boundary, that the segment makes with previous segments. Equivalently, if we define our square to be $[0, 1]^2$ and define the segments of \mathcal{S} accordingly, and if we consider the segment of slope 0 along the x axis to be in \mathcal{S} (and to be present in the square before others are added), then the slice count increases by *exactly* the number of points of intersection with previous segments. Henceforth we adopt this more precise definition of \mathcal{S} and its segments.

We'll make use of the following fact, about the existence in \mathcal{S} of a certain segment that has smaller slope than a given segment.

Useful Fact A. *If a segment $y = Ax - B$ in \mathcal{S} passes through a point p whose x coordinate is a rational number q/r , where $r \leq A$, then the segment $y = (A - r)x - (B - q)$ is in \mathcal{S} and passes through p .*

Proof: Equating y 's and solving for x shows that the two lines in question meet at a point whose x coordinate is q/r , and therefore at p itself: That $y = (A - r)x - (B - q)$ is a segment in \mathcal{S} follows, because $0 \leq A - r < A$ (and because the segment passes through the square). ■

There's an analogous fact for the existence in \mathcal{S} of a segment of larger slope.

Useful Fact B. *If a segment $y = Ax - B$ in \mathcal{S} passes through a point p whose x coordinate is a rational number q/r , where $r \leq N - A$, then the segment $y = (A + r)x - (B + q)$ is in \mathcal{S} and passes through p .*

Proof: Again, the two lines in question meet at p . And the segment $y = (A + r)x - (B + q)$ is in \mathcal{S} , because $0 < A + r \leq N$. ■

So look at what happens to the slice count when we add the A segments of slope A , each of the form $y = Ax - B$, for $0 \leq B < A$. We need to show that these new segments collectively form $\Phi(A)$ intersection points with the old segments, which have smaller slope, and we do this by examining possible x coordinates. Let q/r be the x coordinate, in reduced form, of the intersection of a new segment, say $y = Ax - B$, with an old segment, say $y = ax - b$. (There may be more than one such old segment.) Since

$$\frac{q}{r} = \frac{B - b}{A - a},$$

we have $1 \leq r \leq A$. Furthermore, corresponding to this x coordinate q/r there is among the segments of slope A a unique y coordinate (it is $Aq/r \bmod 1$) and a unique B (it is $\lfloor Aq/r \rfloor$), and therefore just one such intersection point. Conversely,

for any reduced fraction q/r in the interval $[0, .1)$, with $r \leq A$, there exists an old segment whose intersection point with the uniquely determined new segment has x coordinate q/r , by Useful Fact A. So there's a one-to-one correspondence between intersection points and reduced fractions q/r in $[0, .1)$ having $r \leq A$, of which there are $\sum_{r=1}^A \varphi(r) = \Phi(A)$, completing the proof of Equation (2). We defer the asymptotics until Section 7. Incidentally, that set of reduced fractions, together with $1/1$, constitutes what is called the Farey series of order A ; Hardy and Wright [5] devote a full chapter to Farey series.

4. SLICE SHAPES. This section proves several properties of the slices, among them, that only triangles and quadrilaterals occur, and that a quadrilateral's four vertices have at most three distinct x coordinates. We start with the first of the two.

Let L be the lower-left vertex of a slice (all slopes are nonnegative, hence every slice has a vertex that's both lowest and leftmost) and let R be its upper-right vertex, as shown in FIGURE 3. Proceed counterclockwise from L to R around the

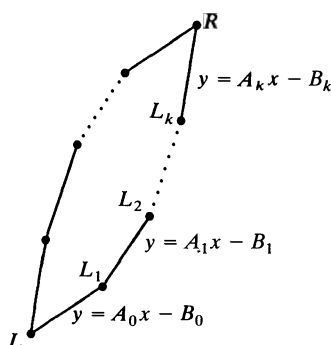


Figure 3. A general slice.

lower boundary of the slice, letting the intermediate vertices encountered be L_1, \dots, L_k , and letting the segments encountered (between consecutive vertices) be $y = A_i x - B_i$, for $i = 0, \dots, k$. Notice that the maximum slope of a segment through L_i is A_i , and the minimum slope is A_{i-1} . We'll see that the sequence of vertices encountered is either L, R or L, L_1, R , so that there are at most two sides on the lower boundary. Moreover, since a rotation by 180° is a symmetry of the sliced-up square, the upper boundary, too, must have at most two sides, hence the whole slice has at most four.

First we handle the slice whose lower boundary includes a side along the line $x = 1$, which isn't of the form $y = Ax - B$. But that slice is easy—it has *exactly* two sides on its lower boundary (when $N = 0$ the slice is the entire square, otherwise it's a triangle).

For any other slice, focus on an intermediate vertex L_i . Its x coordinate is q_i/r_i , not necessarily in reduced form, where $q_i = B_i - B_{i-1}$ and $r_i = A_i - A_{i-1}$. Now suppose that the smaller slope is pretty big: $A_{i-1} \geq N/2$. Then $r_i \leq N - N/2 = N/2$, hence $r_i \leq A_{i-1}$. But by Useful Fact A there's a segment of slope $A_{i-1} - r_i$ through L_i , contradicting the minimality of A_{i-1} . Thus $A_{i-1} < N/2$, giving an upper bound on the smaller slope. Similarly, suppose that the larger slope is pretty small: $A_i \leq N/2$. Then $r_i \leq N/2$, hence $r_i \leq N - A_i$, and by Useful Fact B there's a segment of slope $A_i + r_i$ through L_i , contradicting the maximality

of A_i . Thus $A_i > N/2$, giving a lower bound on the larger slope. Therefore L_2 can't exist, otherwise we'd have both $A_1 < N/2$ and $A_1 > N/2$. So the lower boundary has at most two sides and the whole slice has at most four, as claimed.

That proof also points out a few other properties of every quadrilateral: Each of the four vertices meets one side of small slope—less than $N/2$ —and one side of large slope—greater than $N/2$; each small-slope side is clockwise from an intermediate vertex; each large-slope side is counterclockwise; the two sides at an intermediate vertex differ in slope by more than $N/3$ (otherwise Useful Fact A or B would produce a segment passing through the quadrilateral), and the larger slope is more than twice the smaller; and there is no side in the quadrilateral of slope exactly $N/2$.

Finally, here's a proof that the vertices of a quadrilateral have at most three distinct x coordinates. Define an *aberrant* slice to be one having a vertical side. For the aberrant quadrilateral, which occurs for $N = 0$, there are just two distinct x coordinates.

For a non-aberrant quadrilateral, as in FIGURE 4, let the upper-boundary slopes be $a_0 > a_1$, let the lower-boundary slopes be $A_0 < A_1$, let the upper-boundary intermediate vertex have x coordinate q/r , where $r = a_0 - a_1$, and let the lower-boundary vertices, counterclockwise, have x coordinates x_0, x_1 , and x_2 . Rotating by 180° if necessary, we may assume that $a_0 - a_1 \leq A_1 - A_0$. Now if $x_1 < q/r < x_2$, as in FIGURE 4, then by Useful Fact A there is a segment of slope $A_1 - r$, where $A_0 \leq A_1 - r < A_1$, passing through a point p on the side having slope A_1 ; but that would endow our quadrilateral with a new side. Similarly, if $x_0 < q/r < x_1$ then by Useful Fact B there is a segment of slope $A_0 + r$, where $A_0 < A_0 + r \leq A_1$, passing through the side having slope A_0 , again giving a contradiction. Thus $q/r = x_1$, and the quadrilateral has just three x coordinates, as claimed.

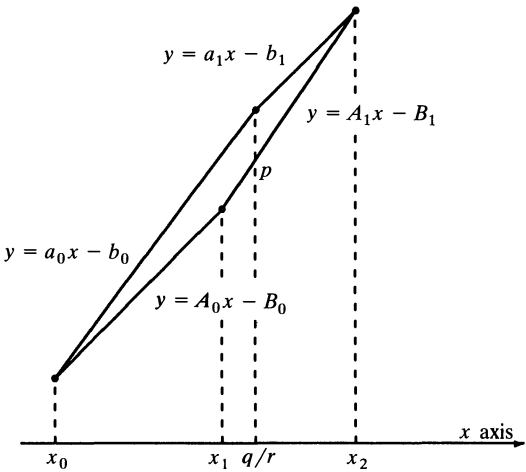


Figure 4. A non-aberrant quadrilateral.

5. COUNTING TRIANGLES. This section counts the number of triangles $T(N)$, showing that

$$T(N) = 2 + 2 \sum_{1 \leq k \leq N} \Phi(k/2), \quad \text{for } N \geq 1. \tag{3}$$

(Notice that by definition (1), the value $\Phi(k/2)$ is perfectly well defined, even for odd k .)

How can the addition of a segment of largest slope divide a slice?

- If the slice is a triangle, the segment either passes through two sides, producing a triangle and a quadrilateral (FIGURE 5A), or through a vertex and a side, producing two triangles (FIGURE 5B).
- If the slice is the aberrant quadrilateral (the entire square), the segment produces two triangles (FIGURE 5C).

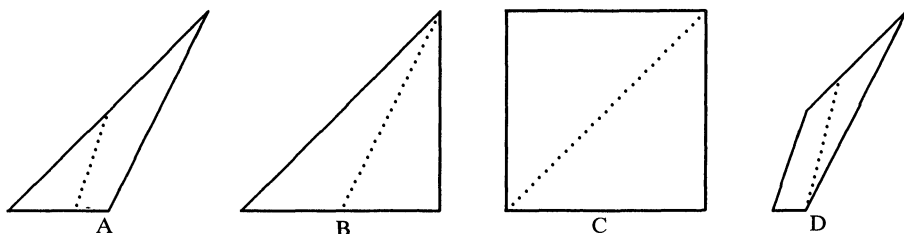


Figure 5. How a largest-slope segment (dotted) can divide a slice.

- If the slice is a non-aberrant quadrilateral, the three- x -coordinate property severely restricts the segment: It must pass through one of the vertices with repeated x coordinate, as well as one of the sides, producing a triangle and a quadrilateral (FIGURE 5D). Passing through two vertices is impossible from its slope, and passing through two sides would produce either a pentagon or a quadrilateral with four distinct x coordinates.

So the triangle-count increase for the segments of slope N equals the number of times one of these new segments passes through a vertex of an old slice that it splits. How many such vertex/slice occurrences are there? Consider the set of previous intersection points (between two or more old segments) through which the segments of slope N pass. Notice that each such point in the interior of the square accounts for two vertex/slice occurrences—one each for the old slices below and to the left, and above and to the right. The same accounting, however, applies to the previous intersection points on the square's boundary (they are, using our convention that $[0, 1]^2$ is the unit square, all on the line $y = 0$): Corresponding to a previous intersection point $(B/N, 0)$, where $1 \leq B < N$, is an old slice above and to the right, which is split by the segment $y = Nx - B$, as well as an old slice below and to the left of the point $(B/N, 1)$ —a point that's not counted, according to our convention—split by the segment $y = Nx - (B - 1)$. And corresponding to the previous intersection point $(0, 0)$, for $N \geq 2$, are two aberrant old slices—one with a side on the y -axis that's above and to the right of $(0, 0)$, and one with a side on the line $x = 1$ that's below and to the left of $(1, 1)$, another point not counted, according to our definition. Thus for $N \geq 2$ the triangle-count increase is twice $P(N)$, where $P(N)$ is the number of previous intersection points through which the segments of slope N pass. And for $N = 1$ there are no previous intersection points, so $P(1) = 0$, but the triangle-count increase is two (the two aberrant triangles formed by the line $y = x$). Thus the

total number of triangles satisfies

$$T(N) = 2 + 2 \sum_{1 \leq k \leq N} P(k), \quad \text{for } N \geq 1, \quad (4)$$

and we're left with determining $P(N)$.

When is a point on a new segment a previous intersection point? Suppose that a new segment $y = Nx - B$ and two old segments $y = a_i x - b_i$, for $i = 1, 2$, have a mutual intersection point. Its x coordinate is rational, say q/r in reduced form. Now $c_i r = N - a_i$, for some positive integer c_i , since q/r is just $(B - b_i)/(N - a_i)$ in reduced form. Furthermore $c_1 \neq c_2$, since $a_1 \neq a_2$. So $\max(c_1, c_2) \geq 2$, hence $r \leq N/2$. Conversely, if segment $y = Nx - B$ passes through a point p whose x coordinate is a rational q/r , with $r \leq N/2$, then both $N - r$ and $N - 2r$ are nonnegative. So by Useful Fact A the segments $y = (N - r)x - (B - q)$ and $y = (N - 2r)x - (B - 2q)$ are also in \mathcal{S} and pass through p , hence p is a previous intersection point. Thus we have the answer: A point on a new segment is a previous intersection point if and only if its x coordinate is a rational number that, in reduced form, is q/r with $r \leq N/2$. Since there are $\Phi(N/2)$ such points, $P(N) = \Phi(N/2)$, and equation (4) implies formula (3) for the number of triangles $T(N)$. That in turn determines the number of quadrilaterals $Q(N)$, which is $S(N) - T(N)$:

$$Q(N) = -1 + \sum_{1 \leq k \leq N} (\Phi(k) - 2\Phi(k/2)), \quad \text{for } N \geq 1. \quad (5)$$

6. THE SURFACE OF CHOICE—A TORUS. Previous discussion was somewhat messy, with all the aberrations and boundaries and boundary points, suggesting that things would be cleaner without the boundaries. We eliminate them by using a torus, as outlined below.

Curl the square into a cylinder, so that the square's top edge, along the line $y = 1$, and its bottom edge, along the line $y = 0$, coincide, and so that the square's left and right edges, along the lines $x = 0$ and $x = 1$, form circles at the ends of the cylinder. Now wrap the cylinder around, joining ends, so that those two circles coincide. The resulting surface is our torus.

The two aberrant triangles of our square, for $N \geq 1$, have now become a single quadrilateral; but that's the only difference in slice counts and "shapes" between the two surfaces. Thus for $N \geq 1$ the additive constants in the expressions for the number of triangles (3), quadrilaterals (5), and total slices (2) disappear.

Moreover, looking at the problem on the torus reveals an interesting phenomenon. Suppose that, before wrapping the cylinder, we first give one end N_0 full twists ($2\pi N_0$ radians). Now when we join ends it's as if we had used lines with slopes between N_0 and $N_0 + N$ instead of between 0 and N —we get the same slice counts in both cases. (The analogue in the plane is to use a parallelogram of unit area whose left edge is still the interval $[0, 1)$ along the y axis but whose right edge is now the interval $[N_0, N_0 + 1)$ along the line $x = 1$.)

Mathematically the torus is a more natural surface for our problem than the square; it is a more difficult surface conceptually, however, which is why our work has been in the square.

7. ASYMPTOTICS. Finally we take a peek at the asymptotics of the quantities involved. A good estimate for $S(N)$ follows easily from one for $\Phi(N)$, whose long

but somewhat confused and controversial history is summarized in the next four paragraphs.

In 1849 Dirichlet [2] determined the leading term, showing that $\Phi(N) = 3N^2/\pi^2 + O(N^{c+\varepsilon})$, where, in terms of the inverse of Riemann's zeta function, c is $\zeta^{-1}(2) \approx 1.73$, and where ε is any positive constant. (The literature sometimes misstates Dirichlet's error term.) In 1874 Franz Mertens [10] improved the error term, defined by

$$E_\Phi(N) = \Phi(N) - 3N^2/\pi^2, \quad (6)$$

from Dirichlet's $O(N^{1.7286\dots})$ to $O(N \log N)$; the proof is pretty easy and appears in Hardy and Wright [5, page 268] and Graham et al. [4, Section 9.3, Problem 6].

Those two 19th-century estimates gave upper bounds on the error. S. S. Pillai and S. D. Chowla [12] in 1930 gave a super-linear lower bound, showing that $E_\Phi(N)$ is not $o(N \log \log \log N)$; that is, there exists a positive constant c such that $|E_\Phi(N)| > cN \log \log \log N$ for infinitely many values of N . Along similar lines, Paul Erdős and Harold N. Shapiro [3] in 1951 showed that there exists a positive constant c such that both $E_\Phi(N) > cN \log \log \log \log N$ for infinitely many values of N and $E_\Phi(N) < -cN \log \log \log \log N$ for infinitely many values of N . (The error E_Φ was once conjectured to be positive for all N .)

So the question, in essence, became how many logs to throw onto the error. In 1953 Arnold Walfisz [19] lowered the upper bound, showing that $E_\Phi(N) = O(N(\log N)^{3/4}(\log \log N)^2)$. He improved that bound further, reducing the pair of exponents from $(\frac{3}{4}, 2)$ to $(\frac{3}{4}, \frac{3}{2})$ in a 1958 paper [20], and then to $(\frac{2}{3}, \frac{4}{3})$ in a 1962 book [21, Chapter 4]. But—and this is the controversy—the book neglected to mention that in 1958 N. M. Korobov [7], improved on Walfisz's 1958 result, obtaining the exponents $(\frac{5}{7}, \gamma)$, for some positive constant γ ; or that in 1960 A. I. Saltykov [13] obtained $(\frac{2}{3}, 1 + \varepsilon)$, showing that

$$\Phi(N) = \frac{3N^2}{\pi^2} + O(N(\log N)^{2/3}(\log \log N)^{1+\varepsilon}), \quad \text{all constants } \varepsilon > 0. \quad (7)$$

We may never know why Walfisz, who died a month after he finished his book, omitted the Korobov and Saltykov results; perhaps he was unaware of them. But there was, at the time, a full-fledged dispute on a related topic. It was between, on the one hand, Korobov [8] and I. M. Vinogradov [17], who claimed without proof that a certain region of the ζ function was zero free, and on the other hand, Walfisz, who said in the notes to Chapter 5 of his book that neither he nor Hans-Egon Richert could verify the Korobov/Vinogradov claim. And nobody else has been able to verify it. Despite requests like A. E. Ingham's [6] for Korobov and Vinogradov to either prove or retract their claim, they apparently have not done so. Korobov's recent book [9] doesn't address the issue, and Vinogradov's book [18] repeats without proof the unsubstantiated claim. Perhaps this related dispute led Walfisz to omit mention of the Korobov and/or Saltykov bounds on $E_\Phi(N)$; neither bound, however, used Korobov's unproven claim (even though Saltykov, who seems to have been a student of Korobov, did use *proved* results from Korobov's paper [8]). Hence notwithstanding the prevalence in the literature of Walfisz's 1962 result, Saltykov's upper bound (7) on E_Φ is undisputed and is the best to date.

The best lower bound is due to Hugh L. Montgomery [11], who in 1987 replaced Erdős and Shapiro's $\log \log \log \log N$ factors with $(\log \log N)^{1/2}$. Thus there remains a gap between these two best bounds, which in essence are $N(\log \log N)^{1/2}$ and $N(\log N)^{2/3}(\log \log N)^{1+\varepsilon}$. And that's a brief history of Φ .

For our purposes here, any of the upper bounds mentioned above will do. Summing Mertens's version, $\Phi(N) = 3N^2/\pi^2 + O(N \log N)$, gives

$$\begin{aligned} S(N) &= 1 + \sum_{1 \leq k \leq N} \Phi(k) \\ &= \frac{3}{\pi^2} \sum_{1 \leq k \leq N} k^2 + O(N^2 \log N) \\ &= \frac{N^3}{\pi^2} + O(N^2 \log N). \end{aligned} \quad (8)$$

So the number of slices is asymptotic to N^3/π^2 , as claimed. Furthermore Equations (3) and (8) show that the number of triangles is

$$\begin{aligned} T(N) &= 2 + 2 \sum_{1 \leq k \leq N} \Phi(k/2) \\ &= 4 \sum_{1 \leq k \leq N/2} \Phi(k) + O(\Phi(N/2)) \\ &= \frac{N^3}{2\pi^2} + O(N^2 \log N). \end{aligned}$$

Thus asymptotically the number of triangles, and therefore the number of quadrilaterals, is half the number of slices. That is,

$$S(N) \sim \frac{N^3}{\pi^2}, \quad T(N) \sim \frac{1}{2}S(N), \quad Q(N) \sim \frac{1}{2}S(N). \quad (9)$$

Incidentally, for small positive N there are more triangles $T(N)$ than quadrilaterals $Q(N)$, and we might conjecture that that's true for all positive N . But it's not; $Q(N)$ first exceeds $T(N)$ at $N = 33$.

What's curious about the number of slices, asymptotically, is the accuracy of the approximation N^3/π^2 . We won't do it here, but it's not hard to remove the log factor from estimate (8) to give

$$S(N) = \frac{N^3}{\pi^2} + O(N^2). \quad (10)$$

Things get really interesting, however, if we restate our problem so that the slopes are strictly less than N , rather than less than or equal to N . While we're at it we'll eliminate the '1 + ' of Equation (2) by using the torus, so that in this new version the number of regions $R(N)$, which is $S(N - 1) - 1$, becomes

$$R(N) = \sum_{1 \leq k < N} \Phi(k), \quad \text{for } N \geq 2, \quad (11)$$

when expressed in terms of Φ . Since this equals $S(N) - \Phi(N) - 1$, and since $\Phi(N) = O(N^2)$, estimate (10) implies $R(N) = N^3/\pi^2 + O(N^2)$. In fact the Prime Number Theorem affords a slightly better error term:

$$R(N) = \frac{N^3}{\pi^2} + o(N^2). \quad (12)$$

That is, the error $E_R(N)$ defined by

$$E_R(N) = R(N) - \frac{N^3}{\pi^2} \quad (13)$$

is $o(N^2)$. The sum $R(N)$ itself seems not to have appeared in the literature, but its error $E_R(N)$ has appeared, in the study of the average size of Φ 's error E_Φ , in the form $\sum_{k=1}^N E_\Phi(k)$. An easy manipulation, summing equation (6) and using equations (11) and (13), shows that

$$E_R(N) = \sum_{1 \leq k < N} E_\Phi(k) - \frac{3N^2}{2\pi^2} + \frac{N}{2\pi^2}. \quad (14)$$

The bound $E_R(N) = o(N^2)$ then follows from Pillai and Chowla's result [12] that $\sum_{k=1}^N E_\Phi(k)$ is asymptotic to $3N^2/(2\pi^2)$. A slightly improved bound, $E_R(N) = O(N^2/\exp(A(\ln N)^{3/5}(\ln \ln N)^{-1/5}))$ for some positive constant A , follows similarly from D. Suryanarayana and R. Sitaramachandra Rao [16], whose estimate comes from Richert and Walfisz's estimate [21, page 191], due primarily to Richert, of a sum involving the Möbius function. And if the Korobov/Vinogradov claim were verified, the improved bound on E_R (along with an analogous bound on the error term in the Prime Number Theorem) would be sharpened by the elimination of the $(\ln \ln N)^{-1/5}$ factor. Even the sharper bound, though, would be quite close to N^2 .

Numerical evidence, on the other hand, suggests that $E_R(N)$ is much smaller than N^2 . For example when $N = 100$ the number of regions $R(N)$ is 101315, and $N^3/\pi^2 \approx 101321$, so $E_R(N)$ is about -6 ; this is somewhat surprising, since N^2 here is 10,000, and the last term added to the sum, $\Phi(99)$, is 3004. The evidence suggests that $E_R(N)$, which is sometimes positive and sometimes negative, is bounded by something closer to $N^{3/2}$. In fact Suryanarayana [15] shows that, assuming the Riemann hypothesis, $\sum_{k=1}^N E_\Phi(k) = 3N^2/(2\pi^2) + O(N^{3/2+\varepsilon})$ for all positive constants ε , and Paolo Codecà [1] proves that this bound holds *if and only if* the Riemann hypothesis is true. Equation (14), along with even Mertens's $O(N \log N)$ bound on $E_\Phi(N)$, therefore implies that $E_R(N)$ is $O(N^{3/2+\varepsilon})$ —which is to say that

$$R(N) = \frac{N^3}{\pi^2} + O(N^{3/2+\varepsilon}), \quad \text{for all constants } \varepsilon > 0$$

—if and only if the Riemann hypothesis is true. So, despite the fairly large gap between essentially N^2 and essentially $N^{3/2}$, further improvements in the error bound await progress on the Riemann hypothesis.

REFERENCES

1. Paolo Codecà. A note on Euler's φ -function. *Arkiv för matematik*, 19:261–263, 1981.
2. [G.] Lejeune Dirichlet. Über die Bestimmung der mittleren Werthe in der Zahlentheorie. *Mathematische Abhandlungen der Königlich Akademie der Wissenschaften zu Berlin*, pages 69–83, published 1851, paper presented 9 August 1849. The cited result is on pages 77–81. Reprinted in his *Werke*, volume 2, pages 49–66 (60–64); Chelsea published both volumes as one in 1969 (vol. 2 was originally published in 1897).
3. Paul Erdős and Harold N. Shapiro. On the changes of sign of a certain error function. *Canadian Journal of Mathematics*, 3:375–385, 1951.
4. Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete Mathematics: A Foundation for Computer Science*. Addison-Wesley, second edition, 1994.
5. G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*. Oxford University Press, fifth edition, 1979.
6. A. E. Ingham. Review 3954. *Mathematical Reviews*, 28(5):764–765, 1964. The request appears in the final paragraph.
7. N. M. Korobov. New number-theoretic estimates. *Doklady Akademii Nauk SSSR*, 119(3):433–434, 1958. In Russian. English summary in *Mathematical Reviews* 20:6395.

8. N. M. Korobov. Estimates of trigonometric sums and their applications. *Uspekhi Matematicheskikh Nauk*, 13, number 4(82), pages 185–192, 1958. In Russian. English summary in *Mathematical Reviews* 21:4939.
9. N. M. Korobov. *Exponential Sums and their Applications*. Kluwer, 1992. A translation by Yu. N. Shakhov from the 1989 Russian original.
10. F[ranz] Mertens. Ueber einige asymptotische Gesetze der Zahlentheorie. *Journal für die reine und angewandte Mathematik*, 77:289–338, 1874. The proof appears on pages 289–291.
11. Hugh L. Montgomery. Fluctuations in the mean of Euler's phi function. *Proceedings of the Indian Academy of Sciences (Mathematical Sciences)*, 97:239–245, 1987.
12. S. S. Pillai and S. D. Chowla. On the error terms in some asymptotic formulae in the theory of numbers (I). *Journal of The London Mathematical Society*, 5:95–101, 1930.
13. A. I. Saltykov. On Euler's function. *Vestnik Moskovskogo Universiteta, Seriya I: Matematika, Mekhanika*, no volume, number 6, pages 34–50, 1960. In Russian. English summary, page 50, or *Mathematical Reviews* 23:A2395.
14. J[acob] Steiner. Einige Gesetze über die Theilung der Ebene und des Raumes. *Journal für die reine und angewandte Mathematik*, 1:349–364, 1826.
15. D. Suryanarayana. On the average order of the function $E(x) = \sum_{n \leq x} \phi(n) - 3x^2/\pi^2$ (II). *Journal of the Indian Mathematical Society*, 42:179–195, 1978. Equation (3.41) gives the stated result.
16. D. Suryanarayana and R. Sitaramachandra Rao. On the average order of the function $E(x) = \sum_{n \leq x} \phi(n) - 3x^2/\pi^2$. *Arkiv för matematik*, 10:99–106, 1972. Equation (1.11) is the result used.
17. I. M. Vinogradov. A new estimate of the function $\zeta(1 + it)$. *Izvestiya Akademii Nauk SSSR, Seriya Matematicheskaya*, 22(2):161–164, 1958. In Russian. English summary in *Mathematical Reviews* 21:2624.
18. I. M. Vinogradov. *Trigonometrical Sums in Number Theory*, pages 9 and 44. Statistical Publishing Society, 204/1 B. T. Road, Calcutta 700 035, 1975. Apparently a translation of the 1971 Russian original.
19. A[rnold] Z. Walfisz. On Euler's function. In *American Mathematical Society Translations*, vol. 4 of series 2, pages 1–29, 1956. English translation, by William H. Simons. The original Russian version is cited as *Akad. Nauk Gruzin. SSSR. Trudy Tbiliss Mat. Inst. Razmadze*, 19:1–31, 1953.
20. A[rnold] Walfisz. Über die Wirksamkeit einiger Abschätzungen trigonometrischer Summen. *Acta Arithmetica*, 4:108–180, 1958. Equation (51), page 115, is the result cited.
21. Arnold Walfisz. *Weylsche Exponentialsummen in der neueren Zahlentheorie*. VEB Deutscher Verlag der Wissenschaften, 1963. A reprint, with his daughter Anna's corrections, of his 1962 book.

Center for Communications Research
 4320 Westerra Court
 San Diego, CA 92121
 op@ccrwest.org

Who Was the Author?

Note on substitution groups of eight letters, Bull. New York Math. Soc., 1894.

Remarks on substitution groups, Amer. Math. Monthly, 1895.

On the commutator groups, Bull. AMS 1898.

On the perfect groups, Amer. J. Math., 1898.

What is group theory? Popular Science Monthly, 1904.

Answer on page 342.

Rational Periodic Points of the Quadratic Function $Q_c(x) = x^2 + c$

Ralph Walde and Paula Russo

1. INTRODUCTION. The family of quadratic maps, $Q_c(x) = x^2 + c$, plays a fundamental role in the understanding of one-dimensional dynamics over both the real and complex numbers. One reason is that virtually every sort of periodic and chaotic behavior is exhibited by at least one member of this 1-parameter family of maps. Also the fixed points and period two points of Q_c can be explicitly computed in terms of the parameter c , making these functions a good source of examples for introductory texts on the subject. Beyond period two, however, explicit calculations become more difficult, and so numerical methods are used to find *approximate* values of attracting periodic points of Q_c . For students being introduced to the material for the first time, examples with only approximate values for periodic points are less satisfying than those which display the precise values.

In this paper we examine some of the periodic orbits of Q_c that can be precisely computed, those which consist entirely of rational numbers. We obtain results that describe explicitly which of the Q_c have both rational fixed points and rational period two points as well as those which have rational period three orbits. These results are particularly appealing as they can be readily used to generate numerous explicit examples. We also show that if c is a rational number with an odd denominator then Q_c has no rational orbits of period greater than 2.

Our theorems have elementary proofs using results from a variety of areas of algebra. Specifically, in addition to the algebraic manipulation of polynomials, we make use of a classical result in number theory on Pythagorean triples and use some elementary results involving p -adic numbers.

2. BACKGROUND FROM DYNAMICS. We present in this section some necessary notation and terminology from dynamical systems. For a more complete introduction to dynamical systems see [3].

For a given function $f(x)$ and an integer $n > 0$, we denote by $f^n(x)$ the composition of f with itself n times. A point λ is said to be a *fixed point* of f if $f(\lambda) = \lambda$ and it is said to be a *period n point* of f if n is the smallest positive integer for which $f^n(\lambda) = \lambda$. In this case the set of points

$$\{\lambda, f(\lambda), f^2(\lambda), f^3(\lambda), \dots, f^{n-1}(\lambda)\}$$

is said to be a *period n orbit* of $f(x)$. When all of the $f^i(\lambda)$ are real we refer to the orbit as a *real orbit*. Similarly, if all of the $f^i(\lambda)$ are rational, we refer to the orbit as a *rational orbit*. We say that λ is a *periodic point* of f if it is a period n point for some positive integer n . Note that authors often use the term prime period n instead of period n for the above property and simply call λ a period n point of f if $f^n(\lambda) = \lambda$ whether n is the smallest such integer or not.

Let λ be a period n point of $f(x)$. We say that λ is *attracting* if

$$|(f^n)'(\lambda)| < 1. \quad (1)$$

In this case there is a neighborhood U about λ such that if $x \in U$ and $x \neq \lambda$ then $|f^n(x) - \lambda| < |x - \lambda|$. Analogously, λ is said to be *repelling* if

$$|(f^n)'(\lambda)| > 1. \quad (2)$$

We shall restrict our attention to the quadratic family of maps defined by $Q_c(x) = x^2 + c$, where c is a complex parameter. For some of our results we will restrict the parameter c to be either a real or a rational number. Note that all rational numbers are assumed to be in lowest terms unless otherwise stated. Occasionally we will consider a function with a fixed value of c and will then simply use the notation $Q(x)$ to refer to this function. Since our main interest is in the periodic points of Q_c , it is helpful to rewrite the definitions of attracting and repelling for this family of functions. Using $Q'_c(x) = 2x$, equations (1) and (2) above become,

$$|2^n \lambda_0 \lambda_1 \dots \lambda_{n-1}| < 1, \quad \text{and} \quad |2^n \lambda_0 \lambda_1 \dots \lambda_{n-1}| > 1. \quad (3)$$

where $\lambda_0 = \lambda$ and $\lambda_k = Q_c^k(\lambda)$ for $k = 1, 2, \dots, n-1$.

3. RATIONAL FIXED POINTS AND RATIONAL PERIOD TWO POINTS OF Q_c .

We begin with a description of the complex, real and rational points of periods one and two for $Q_c(x) = x^2 + c$. The characterization of these points is well known and follows directly by solving the formulas defining the periodic points. However, the usual approach is to fix a value of the parameter c , and then find a formula for the fixed and period two points. Our point of view is to begin with a particular number λ and determine those values of c for which Q_c has λ as a fixed point or period two point. This will allow us to parametrize those quadratic functions that have real or rational points of period one or two.

If λ is a fixed point of Q_c then $Q_c(\lambda) = \lambda^2 + c = \lambda$ implies that $c = \lambda - \lambda^2$ and so $1 - \lambda$ is also a fixed point of Q_c . Clearly, these fixed points are rational numbers if and only if λ itself is a rational number. If μ is a period two point of Q_c , then

$$\frac{(Q_c(Q_c(\mu)) - \mu)}{(Q_c(\mu) - \mu)} = \mu^2 + \mu + c + 1 = 0$$

implies $c = -\mu^2 - \mu - 1$ and so $Q_c(\mu) = -1 - \mu$ is the other period two point. These period two points will both be rational if and only if μ is a rational number. We summarize these results in the following theorem.

Theorem 1. 1. For each complex number λ , the quadratic map, Q_c , has fixed points, λ and $1 - \lambda$, if and only if $c = \lambda - \lambda^2$. These will be distinct if $\lambda \neq \frac{1}{2}$. The map from λ to the function $Q_{\lambda - \lambda^2}$ provides a one-to-one correspondence between the real numbers $\lambda \geq \frac{1}{2}$ and those functions Q_c that have real fixed points. Furthermore, these fixed points are rational if and only if λ is rational.

2. For each complex number $\mu \neq -\frac{1}{2}$, the map Q_c has the period two orbit $\{\mu, -1 - \mu\}$ if and only if $c = -1 - \mu - \mu^2$. The map from μ to the function $Q_{-1 - \mu - \mu^2}$ provides a one-to-one correspondence between the real numbers $\mu > -\frac{1}{2}$ and those Q_c that have a period two orbit consisting of two distinct real numbers. Furthermore, these period two points are rational if and only if μ is rational.

Proof: We need only mention that the restrictions $\lambda \geq \frac{1}{2}$ and $\mu > -\frac{1}{2}$ are needed so that the correspondences between λ and μ and the values for the c 's are single valued. The value $\mu = -\frac{1}{2}$ is excluded since in this case $\mu = -1 - \mu$ and hence is fixed and so there is no c for which $-\frac{1}{2}$ is a point of period 2. \square

If one is interested in explicit examples of rational periodic points for quadratic maps then it is very natural to ask the question: “Which quadratic maps have both rational fixed points and rational period two points?” The following theorem completely characterizes such functions.

Theorem 2. *The function Q_c has both two rational fixed points and a rational period two orbit if and only if it is possible to write $c = -\frac{3}{4} - (X/Y)^2$ where X and Y are members of a Pythagorean triple $\{X, Y, Z\}$. That is, X and Y are positive integers and there exists a positive integer Z with $X^2 + Y^2 = Z^2$. In this case the fixed points are $\frac{1}{2} \pm (Z/Y)$ and the period two points are $-\frac{1}{2} \pm (X/Y)$.*

Proof: If Q_c has rational fixed and period two points then the formulas in Theorem 1 imply that $c = -\lambda^2 + \lambda = -\mu^2 - \mu - 1$ for rational numbers λ and μ . Completing the squares, we obtain

$$-(\lambda - \frac{1}{2})^2 + \frac{1}{4} = -(\mu + \frac{1}{2})^2 - \frac{3}{4} \quad (4)$$

or equivalently,

$$(\mu + \frac{1}{2})^2 + 1 = (\lambda - \frac{1}{2})^2.$$

These equations imply that the rational numbers $\mu + \frac{1}{2}$ and $\lambda - \frac{1}{2}$ have the same denominator so we can write

$$\mu + \frac{1}{2} = \frac{X}{Y} \quad \text{and} \quad \lambda - \frac{1}{2} = \frac{Z}{Y}$$

for integers X, Y and Z . Substituting these values into the previous equation we get $(X/Y)^2 + 1 = (Z/Y)^2$, or equivalently, $X^2 + Y^2 = Z^2$ which is the usual formula for Pythagorean triples. Substituting for μ and λ into our formulas for c we find that $c = -\frac{3}{4} - (X/Y)^2$ and Q_c possesses fixed points $\lambda = \frac{1}{2} \pm (Z/Y)$ and period two points $\mu = -\frac{1}{2} \pm (X/Y)$ as desired. Notice that we can assume that X, Y and Z are all positive. \square

The following corollary gives a somewhat technical characterization of those Q_c possessing both rational fixed and period two points. It can be easily used to generate explicit examples.

Corollary 1. *A complete listing of all examples of Q_c with both rational fixed points and rational period two points is obtained from Theorem 2 and integers $m > n > 0$ with m and n relatively prime and either m or n even by setting either*

$$X = m^2 - n^2, Y = 2mn, \quad \text{and} \quad Z = m^2 + n^2 \quad (5)$$

or

$$X = 2mn, Y = m^2 - n^2, \quad \text{and} \quad Z = m^2 + n^2. \quad (6)$$

Proof: From formula 4 in the proof of Theorem 2 we have

$$c = -\left(\frac{X}{Y}\right)^2 - \frac{3}{4} = -\left(\frac{Z}{Y}\right)^2 + \frac{1}{4}.$$

This shows that it suffices to obtain solutions in which X , Y and Z are positive integers that are pairwise relatively prime, that is, those which are referred to in number theory as primitive Pythagorean triples. The classical solution to the problem of finding such triples is given by equation (5) above when Y is even and equation (6) is obtained when X even and Y odd. (See [6] page 394, for example). \square

Table 1 lists some examples of functions obtained from this corollary with both rational fixed points and rational period two points.

TABLE 1. Some examples of Q_c with both rational fixed and period two points using small values of m and n in the formulas (5) and (6) in Corollary 1.

Formulas	m	n	c	Fixed Points	Period Two Points
(5)	2	1	$-\frac{21}{16}$	$\frac{7}{4}, -\frac{3}{4}$	$\frac{1}{4}, -\frac{5}{4}$
(6)	2	1	$-\frac{91}{36}$	$\frac{13}{6}, -\frac{7}{6}$	$\frac{5}{6}, -\frac{11}{6}$
(5)	3	2	$-\frac{133}{144}$	$\frac{19}{12}, -\frac{7}{12}$	$-\frac{1}{12}, -\frac{11}{12}$
(6)	3	2	$-\frac{651}{100}$	$\frac{31}{10}, -\frac{21}{10}$	$\frac{19}{10}, -\frac{29}{10}$
(5)	4	3	$-\frac{481}{576}$	$\frac{37}{24}, -\frac{13}{24}$	$-\frac{5}{24}, -\frac{19}{24}$
(6)	4	3	$-\frac{2451}{196}$	$\frac{57}{14}, -\frac{43}{14}$	$\frac{41}{14}, -\frac{55}{14}$
(5)	4	1	$-\frac{273}{64}$	$\frac{21}{8}, -\frac{13}{8}$	$\frac{11}{8}, -\frac{19}{8}$
(6)	4	1	$-\frac{931}{900}$	$\frac{49}{30}, -\frac{19}{30}$	$\frac{1}{30}, -\frac{31}{30}$
(5)	5	4	$-\frac{1281}{1600}$	$\frac{61}{40}, -\frac{21}{40}$	$-\frac{11}{40}, -\frac{29}{40}$
(6)	5	4	$-\frac{6643}{324}$	$\frac{91}{18}, -\frac{73}{18}$	$\frac{71}{18}, -\frac{89}{18}$
(5)	5	2	$-\frac{741}{400}$	$\frac{39}{20}, -\frac{19}{20}$	$\frac{11}{20}, -\frac{31}{20}$
(6)	5	2	$-\frac{2923}{1764}$	$\frac{79}{42}, -\frac{37}{42}$	$\frac{19}{42}, -\frac{61}{42}$

4. RATIONAL PERIOD THREE ORBITS OF Q_c . We now wish to consider the existence of rational period three orbits of Q_c . If we were to try to generalize the computations that we used to examine period two points we would note that if ζ is a point of period three of Q_c then

$$Q_c(Q_c(Q_c(\zeta))) = \zeta \quad \text{and} \quad Q_c(\zeta) \neq \zeta.$$

Thus ζ satisfies the equation

$$\frac{(Q_c(Q_c(Q_c(\zeta)))) - \zeta}{Q_c(\zeta) - \zeta} = 0$$

which has the quotient

$$\zeta^6 + \zeta^5 + (3c + 1)\zeta^4 + (2c + 1)\zeta^3 + (3c^2 + 3c + 1)\zeta^2 + (c + 1)^2\zeta + (c^3 + 2c^2 + c + 1) = 0.$$

Unfortunately this sixth degree equation will not be very useful for finding rational period three orbits. It is a cubic equation in c which prevents us from writing c

directly in terms of ζ as we did in Theorem 1. Instead, we give a parametrization of *both* those c with a rational periodic orbit and the period 3 orbit itself in terms of a complex parameter τ . Although the parametrization is a bit unwieldy, it allows us to generate explicit examples of rational period 3 orbits as desired.

Theorem 3. *For any complex number τ with $\tau \neq 0$, $\tau \neq -1$, and $\tau^2 + \tau + 1 \neq 0$, let*

$$c = -\frac{\tau^6 + 2\tau^5 + 4\tau^4 + 8\tau^3 + 9\tau^2 + 4\tau + 1}{4\tau^2(\tau + 1)^2} \quad (7)$$

$$x_1 = \frac{\tau^3 + 2\tau^2 + \tau + 1}{2\tau(\tau + 1)} \quad (8)$$

$$x_2 = \frac{\tau^3 - \tau - 1}{2\tau(\tau + 1)} \quad (9)$$

$$x_3 = -\frac{\tau^3 + 2\tau^2 + 3\tau + 1}{2\tau(\tau + 1)}. \quad (10)$$

It then follows that

1. Q_c has the period three orbit $\{x_1, x_2, x_3\}$. That is, $x_2 = Q_c(x_1)$, $x_3 = Q_c(x_2)$, $x_1 = Q_c(x_3)$, and the three complex numbers x_1 , x_2 , and x_3 , are distinct. Furthermore, every period three orbit of the function Q_c can be described in this manner.
2. The formulas for c , x_1 , x_2 , and x_3 give a one-to-one correspondence between the real numbers $\tau > 0$ and the real period 3 orbits of the functions Q_c .
3. The three points in the orbit are three distinct rational numbers if and only if τ is itself a rational number.

Proof: Given a complex number, τ , subject to the restrictions in the theorem, one can verify that $Q_c(x_1) = x_2$, $Q_c(x_2) = x_3$, and $Q_c(x_3) = x_1$ by straight-forward computations using the above formulas for c and the x_i 's. Note that the condition, $\tau^2 + \tau + 1 \neq 0$ is needed to insure that the x_i 's are distinct. Now, assume ζ is a complex number with $Q_c(Q_c(Q_c(\zeta))) = \zeta$ and assume that $\omega = Q_c(\zeta) = \zeta^2 + c$ is not equal to ζ . We must derive the formulas for c , x_1 , x_2 , and x_3 for some complex τ . We have $c = \omega - \zeta^2$, so that we can write

$$Q_c(x) = x^2 + \omega - \zeta^2.$$

We can now compute

$$\begin{aligned} \zeta &= Q_c(Q_c(\omega)) = Q_c(\omega^2 + \omega - \zeta^2) \\ &= \omega^4 + 2\omega^3 + \omega^2 - 2\omega^2\zeta^2 - 2\omega\zeta^2 + \zeta^4 + \omega - \zeta^2. \end{aligned}$$

Transposing the ζ and factoring we find

$$(\omega - \zeta)(\omega^3 + \omega^2\zeta - \omega\zeta^2 - \zeta^3 + 2\omega^2 + 2\omega\zeta + \omega + \zeta + 1) = 0.$$

Dividing by $\omega - \zeta$ we have an expression equivalent to

$$(\omega + \zeta)^3 + (2 - 2\zeta)(\omega + \zeta)^2 + (1 - 2\zeta)(\omega + \zeta) + 1 = 0.$$

Letting $\tau = \omega + \zeta$, and then solving for ζ we find

$$\zeta = \frac{\tau^3 + 2\tau^2 + \tau + 1}{2\tau(\tau + 1)}$$

which is the formula for x_1 . Now one can derive the other formulas using the equations.

$$\begin{aligned} x_2 &= \omega = \tau - \zeta = \tau - x_1, \\ c &= \omega - \zeta^2 = x_2 - x_1^2, \\ x_3 &= Q_c(x_2). \end{aligned}$$

To prove statement 2 in the theorem we observe that if τ is real then the formulas guarantee that c , x_1 , x_2 and x_3 are real. Conversely, if c , x_1 , x_2 and x_3 are all real then $\tau = x_1 + x_2$ must also be real. To justify making the assumption that $\tau > 0$, use the formulas for the x_i 's to verify that

$$x_1 + x_2 = \tau, \quad x_2 + x_3 = -\frac{\tau + 1}{\tau}, \quad \text{and} \quad x_3 + x_1 = -\frac{1}{\tau + 1}.$$

Using these formulas, it is easy to check that exactly one of these three numbers is positive. In fact, $x_1 + x_2$ is positive if and only if $\tau > 0$, $-1 < \tau < 0$ guarantees that $x_2 + x_3$ is positive, and lastly $x_3 + x_1$ is positive precisely when $\tau < -1$. Thus, by a suitable cyclic renumbering of the x_i 's we can assume that $x_1 + x_2 > 0$. Note that two distinct positive real numbers τ_1 and τ_2 must correspond to distinct period three orbits because of the relationship $\tau = x_1 + x_2$, even though τ_1 and τ_2 might correspond to the same value of c .

Lastly, to prove statement 3 of the theorem, if τ is rational then the formulas guarantee that c , and the x_i 's are rational. Conversely, if c , and the x_i 's are all rational, then $\tau = x_1 + x_2$ must also be rational. \square

The formulas in Theorem 3 for those τ with small numerators and denominators give us examples of rational period 3 orbits with small denominators. Some examples of these are listed in Table 2.

TABLE 2. The period 3 orbit $\{x_1, x_2, x_3\}$ of the functions Q_c using values of τ with small numerators and denominators in the formulas of Theorem 3.

τ	c	x_1	x_2	x_3
1	$-\frac{29}{16}$	$\frac{5}{4}$	$-\frac{1}{4}$	$-\frac{7}{4}$
2	$-\frac{301}{144}$	$\frac{19}{12}$	$\frac{5}{12}$	$-\frac{23}{12}$
$\frac{1}{2}$	$-\frac{421}{144}$	$\frac{17}{12}$	$-\frac{11}{12}$	$-\frac{25}{12}$
3	$-\frac{1849}{576}$	$\frac{49}{24}$	$\frac{23}{24}$	$-\frac{55}{24}$
$\frac{1}{3}$	$-\frac{2689}{576}$	$\frac{43}{24}$	$-\frac{35}{24}$	$-\frac{61}{24}$
$\frac{3}{2}$	$-\frac{6469}{3600}$	$\frac{83}{60}$	$\frac{7}{60}$	$-\frac{107}{60}$
$\frac{2}{3}$	$-\frac{8149}{3600}$	$\frac{77}{60}$	$-\frac{37}{60}$	$-\frac{113}{60}$
$\frac{4}{3}$	$-\frac{49561}{28224}$	$\frac{223}{168}$	$\frac{1}{168}$	$-\frac{295}{168}$

It is of interest to look more carefully at the relationship between c and the values of the points of periods one, two and three. A graph of this relationship can be found in FIGURE 1. It is apparent from this graph just when points of different

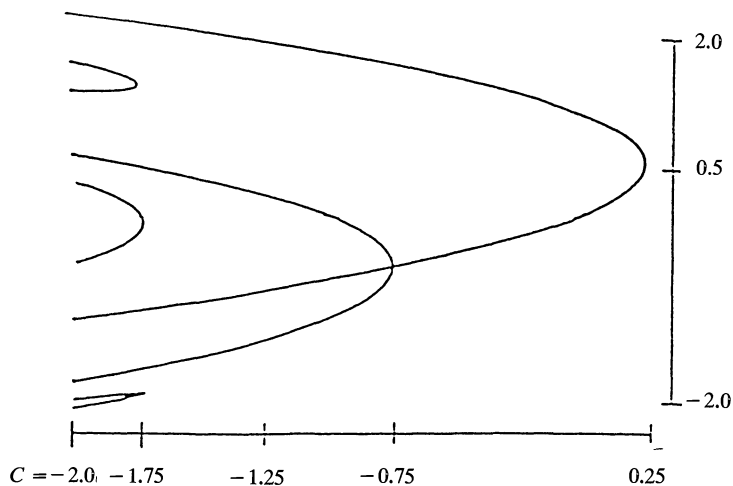


Figure 1. The relationship between c and the values of the points of period 1, 2 and 3.

periods come into existence. Note that this graph displays *all* periodic points of these orders, both attracting and repelling. This is in contrast to the bifurcation diagram for the family, which shows only the *attracting* periodic orbits. A bifurcation diagram drawn to the same scale as the graph in FIGURE 1 can be found in FIGURE 2.

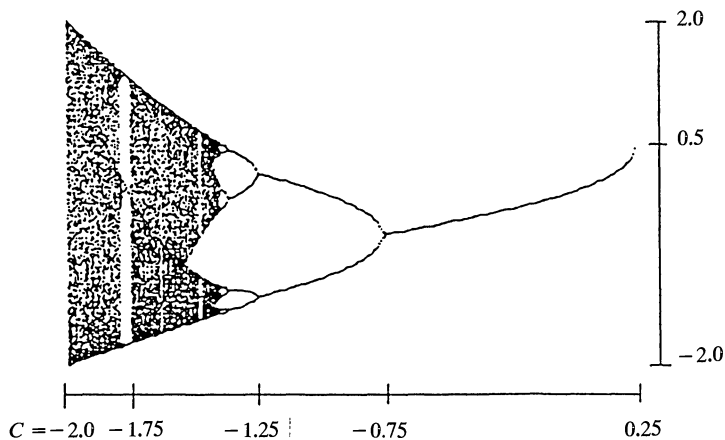


Figure 2. The bifurcation diagram for the family Q_c .

Example 1. In Table 2, the last entry is of particular interest. As $|8x_1x_2x_3| < 1$, this period 3 orbit is attracting. Our computations indicate that this is the attracting rational period 3 orbit having the smallest possible denominator.

Corollary 2. The function Q_c has two period three orbits each with three distinct real numbers if and only if $c < -\frac{7}{4}$. If $c = -\frac{7}{4}$, then Q_c has a single period three orbit consisting of the three distinct irrational roots of $x^3 + \frac{1}{2}x^2 - \frac{9}{4}x - \frac{1}{8} = 0$.

Proof: If one differentiates formula (7) in Theorem 3 and then factors the result, one finds that

$$\frac{dc}{d\tau} = - \frac{(\tau^3 + \tau^2 - 2\tau - 1)(\tau^2 + \tau + 1)^2}{2\tau^3(\tau + 1)^3}$$

so that c possesses a single extremum for $\tau > 0$, a maximum, which occurs at the positive root of $\tau^3 + \tau^2 - 2\tau - 1 = 0$. To find the maximum value of c , one makes repeated use of the substitution $\tau^3 = -\tau^2 + 2\tau + 1$ in the formula for c and discovers that $c = -\frac{7}{4}$. For any value of $c < -\frac{7}{4}$, there are two different positive values of τ yielding the c and thus two different period three orbits of real numbers. If $c = -\frac{7}{4}$, then any complex period three point must be a solution of the equation

$$0 = \frac{Q_c(Q_c(Q_c(x))) - x}{Q_c(x) - x} = \left(x^3 + \frac{1}{2}x^2 - \frac{9}{4}x - \frac{1}{8}\right)^2$$

and hence there is a single orbit of period 3. \square

Corollary 3. *The sets of c for which Q_c has a rational orbit of period 1, 2 and 3 are dense in the sets $c \leq \frac{1}{4}$, $c \leq -\frac{3}{4}$, and $c \leq -\frac{7}{4}$, respectively.*

Proof: This follows directly from the continuity of the functions that describe the parameter c in Theorems 1 and 3. \square

One should note at this point that Theorems 1 and 3 as well as the above corollary have an algebraic geometry flavor. That is, the parameterization of a curve provides a description of the *rational points* of the curve and shows that they are dense in the curve. See [5] for information.

5. BACKGROUND FROM P -ADIC NUMBERS. Up to this point we have focused on giving some partial answers to the question of which Q_c have rational periodic orbits. We now turn to the consideration of which rational values of c give rise to Q_c which *do not* have rational orbits of any period as well as those that have very few. For these results a different algebraic approach is necessary. The main tools used in the proofs are results about the p -adic numbers. We give here some of the necessary background. A full treatment of p -adic numbers can be found in the text [1] or briefer comments concerning them can be found in the recent article [2].

We begin with some of the basic definitions. For any prime number p , the p -adic valuation or norm, $|x|_p$, on the set of rational numbers, can be defined by

$$|0|_p = 0, \quad \text{and} \quad |x|_p = p^{-k}, \quad \text{if } x = p^k \frac{m}{n}$$

where k is the unique integer such that the non-zero integers m and n are not divisible by p .

This definition can be extended to a larger set of numbers called the p -adic numbers. For any prime number p , the set of p -adic numbers consists of all formal series of the form

$$p^k(a_0 + a_1p^1 + a_2p^2 + a_3p^3 + \dots)$$

with k any integer and each a_k an integer with $0 \leq a_k \leq p-1$. The p -adic

valuation $|x|_p$ is then defined on this set of numbers by

$$|0|_p = 0, \quad \text{and} \quad |x|_p = p^{-k}, \quad \text{if } x = p^k(a_0 + a_1p^1 + a_2p^2 + \dots)$$

where k is chosen so that a_0 is positive. Addition, subtraction, multiplication, and division can be defined on the p -adic numbers by using the usual power series rules and then adjusting the coefficients to satisfy $0 \leq a_k \leq p - 1$. With these definitions, the p -adic numbers form a field which is complete with respect to the metric defined by

$$d(x, y) = |x - y|_p.$$

For any prime p , the set of p -adic numbers contains all positive and negative rational numbers expressed as eventually repeating series. For example, in the set of 2-adic numbers,

$$-1 = 1 + 2^1 + 2^2 + 2^3 + 2^4 + \dots$$

because if we add 1 to this series we get the series for 0. Similarly,

$$-\frac{1}{5} = 1 + 2^1 + 2^4 + 2^5 + 2^8 + 2^9 + \dots$$

because multiplying this series by $5 = 1 + 2^2$ we get the series for -1 . Those p -adic series that are not eventually repeating are not rational numbers and, in general, may not correspond to either real or complex numbers. The p -adic valuation on the p -adic numbers has several useful properties which we state in the theorem below.

Theorem 4. *If x and y are p -adic numbers then*

1. $|x|_p \geq 0$ with $|x|_p = 0$ if and only if $x = 0$.
2. $|xy|_p = |x|_p |y|_p$.
3. $|x + y|_p \leq \max(|x|_p, |y|_p)$, with equality holding if $|x|_p \neq |y|_p$.

Some of our results depend on the following special result which can be deduced directly from Theorem 4.4 on page 60 of [1].

Theorem 5. *A non-zero integer b has a square root in the 2-adic numbers if and only if $b = 2^{2n}a$, where a is congruent to 1 modulo 8. ($a \equiv 1 \pmod{8}$) In this case b has two distinct 2-adic square roots.*

Substituting the p -adic valuation for the real or complex absolute value in equation (3) we obtain a definition of attracting and repelling periodic points for the quadratic function thought of as a map from the p -adic numbers to themselves. Thus, for any p -adic number c , a p -adic period n point λ of Q_c is *attracting* if

$$|2^n \lambda_0 \lambda_1 \lambda_2 \dots \lambda_{n-1}|_p < 1$$

and it is *repelling* if

$$|2^n \lambda_0 \lambda_1 \lambda_2 \dots \lambda_{n-1}|_p > 1$$

where the λ_k are defined as in Section 2.

6. PERIODIC POINTS OF Q_c OVER THE p -ADIC NUMBERS. Although we are mainly interested in the dynamic behavior of Q_c as a function over the rational numbers, it will be helpful to establish some results about the periodic points of the quadratic functions over the p -adic numbers. Our first theorem gives a useful relationship between the parameter c and the periodic points of Q_c .

Theorem 6. *If λ is a periodic point of Q_c over the field of p -adic numbers with $|\lambda|_p > 1$ then*

$$|\lambda|_p^2 = |c|_p \quad \text{and} \quad |Q_c(\lambda)|_p = |\lambda|_p.$$

Proof: Assume λ is a periodic point of Q_c with $|\lambda|_p > 1$. We will consider three cases. First, let us assume that $|\lambda|_p^2 > |c|_p$. Then,

$$|Q_c(\lambda)|_p = |\lambda^2 + c|_p = \max(|\lambda|_p^2, |c|_p) = |\lambda|_p^2 > |\lambda|_p.$$

Applying the same argument to $Q_c(\lambda)$ we obtain the increasing sequence,

$$|\lambda|_p < |Q_c(\lambda)|_p < |Q_c(Q_c(\lambda))|_p < |Q_c(Q_c(Q_c(\lambda)))|_p < \dots$$

This contradicts the fact that λ is periodic.

Second, we assume that $|\lambda|_p^2 < |c|_p$. Then

$$|Q_c(\lambda)|_p = |\lambda^2 + c|_p = \max(|\lambda|_p^2, |c|_p) = |c|_p > |\lambda|_p^2 > |\lambda|_p.$$

Now $\mu = Q_c(\lambda)$ has the property that

$$|\mu|_p^2 = |c|_p^2 > |c|_p$$

so $\mu = Q_c(\lambda)$ can be treated as in the first case. Thus, once again

$$|\lambda|_p < |Q_c(\lambda)|_p < |Q_c(Q_c(\lambda))|_p < |Q_c(Q_c(Q_c(\lambda)))|_p < \dots$$

contradicting the fact that λ is periodic.

For the last case we assume that $|\lambda|_p^2 = |c|_p$. Let $\mu = Q_c(\lambda)$ which is also a periodic point of Q_c . By using the first two cases above we can deduce that $|\mu|_p^2 = |c|_p$ so that $|\lambda|_p^2 = |\mu|_p^2$. Since we know that $|\lambda|_p \neq 0$ and $|\mu|_p \neq 0$ we can conclude that $|\lambda|_p = |\mu|_p = |Q_c(\lambda)|_p$ as required. \square

It is clear that if λ is a rational number, then it is a periodic point of Q_c viewed as a function over the p -adic numbers if and only if it is periodic over the real numbers. Thus, we can use Theorem 6 to obtain information about the rational periodic points of Q_c .

Corollary 4. *If j/k is a rational periodic point of Q_c with $c = m/n$ and both j/k and m/n are rational numbers expressed in lowest terms with k and n positive integers then $n = k^2$.*

Proof: For any prime p , Theorem 6 implies that if $|j/k|_p > 1$ then

$$\left| \frac{j}{k} \right|_p^2 = \left| \frac{m}{n} \right|_p.$$

Thus if p^a with $a > 0$ is the largest power of p that divides k then p^{2a} is the largest power of p that divides n and so p divides k if and only if it divides n . Since the largest power of p dividing n is the square of the largest power dividing k for any prime p , we can conclude that $n = k^2$. \square

Corollary 5. *If c is rational, then Q_c has at most finitely many rational periodic points.*

Proof: By the previous corollary we can assume that $c = (m/k^2)$ and each rational periodic point has the form j/k . Thus all of the rational periodic points

have the same denominator. In addition, for any x with $|x| > |c| + 1$, the triangle inequality implies that

$$|Q_c(x)| \geq |x|^2 - |c| > |x|^2 - |x| + 1 = (|x| - 1)^2 + |x| \geq |x|.$$

Thus x cannot be periodic. This implies that the set of all rational periodic points of Q_c is a subset of the finite set

$$\left\{ \frac{j}{k} : \left| \frac{j}{k} \right| \leq |c| + 1 = \frac{|m|}{k^2} + 1 \right\}. \quad \square$$

The following theorem gives a sufficient condition for Q_c to have no 2-adic periodic points other than fixed points.

Theorem 7. *If c is a rational number with an even numerator, then Q_c has two 2-adic fixed points and no other 2-adic periodic points. In fact, one of these fixed points attracts all 2-adic numbers with norms less than 1 and the other fixed point attracts all those with norms equal to 1. (The p -adic numbers of norm greater than 1 will be shown to have unbounded orbits.)*

Proof: The function $Q_c(x) = x^2 + (2m/n)$ has a fixed point whenever the equation $x^2 + 2m/n = x$ has a 2-adic solution. Transposing x to the left side of this equation and completing the square gives us the equation

$$\left(x - \frac{1}{2} \right)^2 = \frac{1}{4} - \frac{2m}{n} = \frac{n^2 - 8mn}{(2n)^2}.$$

Thus Q_c has fixed points if and only if $n^2 - 8mn$ is a square in the 2-adic field. By Theorem 5 it suffices to show that $n^2 - 8mn \equiv 1 \pmod{8}$. This is equivalent to showing that $n^2 \equiv 1 \pmod{8}$ as it is clear that $8mn \equiv 0 \pmod{8}$. Since n is odd we have that

$$n^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 4k(k + 1) + 1$$

and since either k or $k + 1$ must be even, $n^2 \equiv 1 \pmod{8}$. Thus, Q_c has a pair of 2-adic fixed points.

Let α denote one of these 2-adic fixed points. As we saw in Section 3, $1 - \alpha$ is also a fixed point of Q_c and $\alpha(1 - \alpha) = c = (2m/n)$. Now,

$$|\alpha|_2 |1 - \alpha|_2 = |\alpha(1 - \alpha)|_2 = |c|_2 = \left| \frac{2m}{n} \right|_2 \leq \frac{1}{2},$$

where the last inequality follows from the definition of the p -adic norm. Part 3 of Theorem 4 implies that,

$$|\alpha|_2 < 1 \quad \text{and} \quad |1 - \alpha|_2 = \max(1, |\alpha|_2) = 1$$

or

$$|1 - \alpha|_2 < 1 \quad \text{and} \quad |\alpha|_2 = |1 - (1 - \alpha)|_2 = \max(1, |1 - \alpha|_2) = 1.$$

Since α could denote either fixed point we may assume that $|\alpha|_2 < 1$ and $|1 - \alpha|_2 = 1$. We will show that no 2-adic numbers other than α and $1 - \alpha$ can be periodic points of Q_c .

Assume a is a 2-adic number with $a \neq \alpha$. To show that a cannot be a periodic point of Q_c we consider three cases, $|a|_2 < 1$, $|a|_2 = 1$, and $|a|_2 > 1$. We begin

with $|a|_2 > 1$. In this case we have,

$$|a - \alpha|_2 \leq \max(|a|_2, |\alpha|_2) < 1.$$

We may write $|a - \alpha|_2 = 2^{-k}$ for some positive integer k and so $a - \alpha = 2^k u$ for some 2-adic number u with $|u|_2 = 1$. Now,

$$\begin{aligned} |Q_c(a) - \alpha|_2 &= |Q_c(\alpha + 2^k u) - \alpha|_2 \\ &= |\alpha^2 + 2^{k+1}\alpha u + 2^{2k}u^2 + (\alpha - \alpha^2) - \alpha|_2 \\ &= |2^{k+1}(\alpha u + 2^{k-1}u^2)|_2 \leq 2^{-(k+1)} \\ &< |a - \alpha|_2, \end{aligned}$$

which is less than 1. We also have

$$|Q_c(a)|_2 = |a^2 + \alpha(1 - \alpha)| \leq \max(|a^2|_2, |\alpha(1 - \alpha)|_2) < 1$$

which allows the computation to be repeated to obtain the decreasing sequence

$$|a - \alpha|_2 > |Q_c(a) - \alpha|_2 > |Q_c(Q_c(a)) - \alpha|_2 > |Q_c(Q_c(Q_c(a))) - \alpha|_2 > \dots$$

This shows that a cannot be a periodic point of Q_c because it is attracted to α as each term in the above sequence must be a different power of $\frac{1}{2}$.

In a similar fashion, one can show that any a with $|a|_2 = 1$ cannot be a periodic point of Q_c and is in fact attracted to $1 - \alpha$.

Finally, we assume that a is a 2-adic number $|a|_2 > 1$. We find that

$$|Q_c(a)|_2 = \left| a^2 + \frac{2m}{n} \right|_2 = \max\left(|a^2|_2, \left| \frac{2m}{n} \right|_2\right) = |a|_2^2 > |a|_2.$$

Continuing the process, we obtain

$$|a|_2 < |Q_c(a)|_2 < |Q_c(Q_c(a))|_2 < \dots$$

which shows that a can not be a periodic point in this case either. Thus, the two fixed points are the only 2-adic numbers that can be periodic points. \square

The following corollary restates this theorem for the real quadratic family.

Corollary 6. *If $Q_c(x) = x^2 + 2m/n$ where m and n are integers and n is odd then Q_c can have no rational periodic points other than fixed points.*

Proof: As previously noted, we can consider Q_c as a function over the 2-adic numbers and so by Theorem 7, Q_c has no periodic points other than fixed points. But the 2-adic numbers contain all rational numbers so that Q_c can have no rational periodic points other than, possibly, fixed points. \square

Example 2. The quadratic $Q_{-2}(x) = x^2 - 2$ is frequently used as an example of a chaotic function over the real numbers. (See Section 4.4 of [3], for example). For any $n \geq 1$, $Q_{-2}(x)$ has 2^n real periodic points of a period dividing n and all of them are repelling. Thus $Q_{-2}(x)$ has an infinite number of real periodic points and these periodic points are dense in the interval $[-2, 2]$. Corollary 8 is one way to show that *none* of these periodic points are rational except the fixed points 2 and -1 .

We now give a sufficient condition for Q_c to have only 2-adic period two points.

Theorem 8. If $c = (m/n)$ with m and n both odd integers then Q_c , considered as a function over the 2-adic numbers has one period 2 orbit and has no other 2-adic periodic points.

Proof: Let m and n be odd integers and assume that $Q_c(x) = x^2 + (m/n)$ has a point of period 2. Then the equation

$$\frac{Q_c(Q_c(x)) - x}{Q_c(x) - x} = x^2 + x + c + 1 = x^2 + x + \frac{m}{n} + 1 = 0$$

must have a solution in the field of 2-adic numbers. Completing the square we obtain

$$\left(x + \frac{1}{2}\right)^2 = -\frac{3}{4} - \frac{m}{n} = \frac{-3n^2 - 4mn}{4n^2},$$

implying that Q_c has period 2 points if and only if $(-3n^2 - 4mn)$ is a square in the 2-adic field. By Theorem 5, we need to show that $-3n^2 - 4mn \equiv 1 \pmod{8}$. But $m = 2s + 1$ and $n = 2r + 1$ are both odd and so

$$\begin{aligned} -3n^2 - 4mn &\equiv -n(3n + 4m) \equiv -(2r + 1)(6r + 8s + 7) \equiv -(2r + 1)(6r + 7) \\ &\equiv -r(12r + 12 + 8) - 7 \equiv -12r(r + 1) - 7 \equiv -7 \\ &\equiv 1 \end{aligned}$$

since $r(r + 1)$ is even. Thus, Q_c has a 2-adic period 2 orbit.

Let β denote one of the 2-adic period 2 points of Q_c . As was shown in section 3, $\gamma = -1 - \beta$ is also a period 2 point of Q_c and $c = -1 - \beta - \beta^2 = (m/n)$. Note that

$$|\beta|_2 |\gamma|_2 = |\beta(-1 - \beta)|_2 = |c + 1|_2 = \left| \frac{m + n}{n} \right|_2 < 1$$

where the last inequality follows from the fact that $m + n$ is even. As in the proof of Theorem 7 we may assume that $|\beta|_2 < 1$ and $|\gamma|_2 = 1$. We will show that no 2-adic numbers other than β and γ can be periodic points of Q_c .

We begin with a 2-adic number, a , with $|a|_2 < 1$ and $a \neq \beta$, so that $|a - \beta|_2 \leq \max(|a|_2, |\beta|_2) < 1$. Proceeding as in Theorem 7, one can show that

$$|Q_c(a) - \gamma|_2 < |a - \beta|_2 < 1.$$

We may interpret the above inequality as saying that if $|a|_2 < 1$ then $Q_c(a)$ is closer to γ than a is to β . In addition, $|\gamma|_2 = 1$ and $|Q_c(a) - \gamma|_2 < 1$ implies that $|Q_c(a)|_2 = 1$. To continue the process, we need a similar computation for any 2-adic a' having the properties that $|a'|_2 = 1$ and $a' \neq \gamma$. As above, one can show that $Q_c(a')$ is closer to β than a' is to γ . Combining these results, we can conclude that if $|a|_2 < 1$ then

$$|a - \beta|_2 > |Q_c(Q_c(a)) - \beta|_2 > |Q_c(Q_c(Q_c(Q_c(a)))) - \beta|_2 > \cdots$$

A similar inequality also holds in the case $|a|_2 = 1$. This shows that if a is a 2-adic number with $|a|_2 \leq 1$ and $a \neq \beta$, $a \neq \gamma$ then a cannot be a periodic point of Q_c . The equations can also be interpreted as saying that the set of all 2-adic numbers a with $|a|_2 \leq 1$ is attracted to the period 2 orbit consisting of β and $-1 - \beta$.

Finally assume that a is a 2-adic number $|a|_2 > 1$ then we find that

$$|Q_c(a)|_2 = \left| a^2 + \frac{m}{n} \right|_2 = |a|_2^2 > |a|_2.$$

Now

$$|a|_2 < |Q_c(a)|_2 < |Q_c(Q_c(a))|_2 < \dots$$

shows that a cannot be a periodic point in this case either. Thus the two period 2 points are the only 2-adic numbers that can be periodic points of Q_c . \square

Corollary 7. *If $Q_c(x) = x^2 + m/n$ where m and n are odd integers then Q_c can have no rational periodic points of any prime period other than 2.*

Proof: The same observations as those in the proof of the corollary to Theorem 7 hold. \square

There are many natural questions that arise from our results. While we cannot guarantee that these are open questions, we list below some questions whose answers are unknown to us.

1. Are there any rational periodic orbits of a quadratic Q_c of period greater than 3? The results for periods 1, 2, and 3 would lead one to suspect that there must be.
2. Over the complex numbers, the quadratic functions Q_c have six period 3 points (except for $c = -\frac{7}{4}$). Theorem 3 shows that for certain rational numbers c , three of these period 3 points are rational numbers. Are there any cases when all six of the period 3 points are rational?
3. Are there any cases when Q_c has a rational period three orbit and either rational fixed points or rational period 2 points?

Note: After our paper was accepted for publication, we discovered three papers that contain some of our results. Corollary 5 follows from a more general result in [7], Theorem 5 appears in [8], and Theorem 3 and a negative answer to question 2 above are given in [9].

REFERENCES

1. Bachman, George *Introduction to p-adic Numbers and Valuation Theory*, Academic Press, 1964.
2. Cuoco, Albert A., *Visualizing the p-adic Integers*, Mathematical Monthly, April 1991, pp. 355–364.
3. Devaney, Robert, *An Introduction to Chaotic Dynamical Systems*, Addison Wesley, 1987.
4. Devaney, Robert, *Chaos, Fractals, and Dynamics: Computer Experiments in Mathematics*, Addison Wesley, 1990.
5. Kendig, Keith, *Elementary Algebraic Geometry*, Springer-Verlag, 1977.
6. Rosen, Kenneth H., *Elementary Number Theory and Its Applications*, Addison Wesley, 1984.
7. Morton, Patrick, *Arithmetic properties of periodic points of quadratic maps*, Acta Arithmetica, LXII.4 (1992) pp. 343–372.
8. Narkiewicz, W., *Polynomial cycles in algebraic number fields*, Colloquium Mathematicum 58 (1989), pp. 151–155.
9. Thiran, E., Verstegen, D., and Weyers, J., *p-adic Dynamics*, Journal of Statistical Physics, Vol. 54, Nos. 3/4, 1989, pp. 893–913.

Department of Engineering
& Computer Science
Trinity College
Hartford, CT 06106
ralph.walde@trincoll.edu

Department of Mathematics
Trinity College
Hartford, CT 06106
paula.russo@trincoll.edu

Fréchet vs. Carathéodory

Ernesto Acosta G. and Cesar Delgado G.

1. INTRODUCTION. When we read Kuhn's paper, The Derivative a la Carathéodory [Kuh], we were very impressed how Carathéodory's formulation of derivative simplifies the proofs of the basic differentiability theorems for real-valued functions of one variable, in particular the proof of the Chain Rule. We saw that the strength of Carathéodory's formulation relies on the concept of continuity and that the proofs strongly use the properties of continuous functions. Another advantage about this formulation is that it does not require the difference quotient, which is the key to generalize it to functions of several variables.

Let us recall the usual definition of differentiability. Let f be a real valued function defined on \mathbb{R} and a be a real number. We say that f is differentiable at a if the limit

$$\lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a}$$

exists. We can say this in another way. Let ϕ be defined by

$$\phi(x) = \frac{f(x) - f(a)}{x - a}.$$

Then, f is differentiable at a if and only if ϕ has a removable discontinuity at a . This is a motivation to state the following characterization of differentiability:

$$\begin{aligned} &f \text{ is differentiable at } a \text{ if there exists a function } \phi \\ &\text{which is continuous at } a \text{ and such that} \end{aligned} \tag{1}$$
$$f(x) - f(a) = \phi(x)(x - a).$$

The latter is Carathéodory's characterization of differentiability [Car]. We will extend it to functions $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ and explore some of the advantages it has over Fréchet's characterization [Fre]. The reader could see without difficulty that all we do in §2, §3, and §4 can be redone replacing \mathbb{R}^n and \mathbb{R}^m by two arbitrary Hilbert spaces, and that Carathéodory's definition makes perfect sense in general linear topological spaces [Aco, Del].

2. THE TWO FORMULATIONS. First we extend definition (1) to functions $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ and second we show it is equivalent to the corresponding one in the Fréchet sense.

If we read (1) thinking of f as a function from \mathbb{R}^n to \mathbb{R}^m we have to think of $x - a$ and $f(x) - f(a)$ as points (or vectors) in \mathbb{R}^n and \mathbb{R}^m respectively. The question is: what is $\phi(x)$? When we multiply $x - a \in \mathbb{R}^n$ by $\phi(x)$ we get $f(x) - f(a) \in \mathbb{R}^m$. Thus we can think of $\phi(x)$ as a $m \times n$ matrix which gives us a good interpretation of (1). So, we consider ϕ as a function on \mathbb{R}^n taking values in the space $M_{m \times n}$ of real matrices.

the space $M_{m \times n}$ of real matrices.

Now we can give the following definition:

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $a \in \mathbb{R}^n$. We say that f is differentiable at a if there exists a function $\phi : \mathbb{R}^n \rightarrow M_{m \times n}$ which is continuous at a and satisfies

$$(2)$$

$$f(x) - f(a) = \phi(x)(x - a).$$

We call such a function ϕ a slope function for f at a .

With respect to the second, let us recall Fréchet's definition [Spi, page 14]:

f is differentiable at a if there exists a linear transformation $\lambda : \mathbb{R}^n \rightarrow \mathbb{R}^m$ such that

$$(3)$$

$$\lim_{x \rightarrow a} \frac{\|f(x) - f(a) - \lambda(x - a)\|}{\|x - a\|} = 0.$$

If f satisfies (2) we will say that f is Carathéodory differentiable at a and if satisfies (3) that f is Fréchet differentiable at a . However we have the following theorem which states the equivalency between (2) and (3).

Theorem 1. *Every Fréchet differentiable function is Carathéodory differentiable and vice versa.*

Proof: We can see without difficulty that (2) implies (3). In fact, if we assume the existence of ϕ , we have

$$\frac{\|f(x) - f(a) - \phi(a)(x - a)\|}{\|x - a\|} = \frac{\|(\phi(x) - \phi(a))(x - a)\|}{\|x - a\|} \leq \|\phi(x) - \phi(a)\|.$$

Due to the continuity of ϕ at a we get (3).

Let us show now that (3) implies (2). Assume that λ exists and define ϕ^1 by

$$\phi(x) = \begin{cases} \frac{1}{\|x - a\|^2} \{(f(x) - f(a) - \lambda(x - a)) \otimes (x - a)\} + \lambda, & x \neq a \\ \lambda, & x = a. \end{cases}$$

We can see immediately from the definition of ϕ that $\phi(x)(x - a) = f(x) - f(a)$.

We have to prove the continuity of ϕ at a . But

$$\|\phi(x) - \phi(a)\| \leq \frac{\|f(x) - f(a) - \lambda(x - a)\|}{\|x - a\|}$$

and since f satisfies (3) we get (2).

3. A UNICITY THEOREM. We know that if λ exists in (3), it is unique [Spi, Theorem 2-1]. λ is called the derivative of f at a and is denoted by $Df(a)$. Unfortunately this uniqueness is not true for ϕ in (2), as we can see in the following example. Let $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ be defined by $f(x, y) = xy$, and pick a point

¹ \otimes represents tensor product. I.e., if $u \in \mathbb{R}^m$ and $v, w \in \mathbb{R}^n$ then $u \otimes v \in M_{m \times n}$ is defined by

$$(u \otimes v)w = (v \cdot w)u$$

where $v \cdot w$ is the inner product of v and w .

$(a, b) \in \mathbb{R}^2$. Then

$$\begin{aligned} f(x, y) - f(a, b) &= (b, x)(x - a, y - b) \\ &= (y, a)(x - a, y - b). \end{aligned}$$

Therefore $\phi(x, y) = (b, x)$ and $\psi(x, y) = (y, a)$ are two different slope functions for f at (a, b) . However observe that $\phi(a, b) = \psi(a, b)$ and we can ask if this is true in general. The answer is yes!

Theorem 2. *If ϕ and ψ are two slope functions for f at a then $\phi(a) = \psi(a)$.*

Proof: Assume that ϕ and ψ are two slope functions for f at a and let $\eta(x) = \phi(x) - \psi(x)$. Then

$$\eta(x)(x - a) = 0,$$

and so

$$\begin{aligned} \|\eta(a)(x - a)\| &= \|(\eta(a) - \eta(x))(x - a)\| \\ &\leq \|\eta(a) - \eta(x)\| \|x - a\|. \end{aligned}$$

Therefore we have that

$$\left\| \eta(a) \left(\frac{x - a}{\|x - a\|} \right) \right\| \leq \|\eta(a) - \eta(x)\|.$$

Since η is continuous at a , we conclude that $\eta(a) = 0$ and then that $\phi(a) = \psi(a)$.

We can talk then about the derivative of f at a following Carathéodory which is precisely $\phi(a)$. Evidently we have $\phi(a) = Df(a)$.

4. BASIC DIFFERENTIATION THEOREMS IN \mathbb{R}^n . The three basic differentiation theorems in \mathbb{R}^n are essentially the one of linearity, the one of the chain rule and the one of critical points. Here definition (2) starts showing its virtues.

Theorem 3. *If $f, g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ are differentiable at $a \in \mathbb{R}^n$ then $\alpha f + \beta g$, $\alpha, \beta \in \mathbb{R}$, is also differentiable at a and*

$$D(\alpha f + \beta g)(a) = \alpha Df(a) + \beta Dg(a).$$

Proof: Let $f, g: \mathbb{R}^n \rightarrow \mathbb{R}^m$ be differentiable at $a \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$. Then

$$\begin{aligned} (\alpha f + \beta g)(x) - (\alpha f + \beta g)(a) &= \alpha(f(x) - f(a)) + \beta(g(x) - g(a)) \\ &= (\alpha\phi + \beta\psi)(x)(x - a), \end{aligned}$$

where ϕ and ψ are slope functions for f and g respectively. Since ϕ and ψ are continuous at a , $\alpha\phi + \beta\psi$ is also continuous at a . Therefore $\alpha f + \beta g$ is differentiable at a and

$$\begin{aligned} D(\alpha f + \beta g)(a) &= \alpha\phi(a) + \beta\psi(a) \\ &= \alpha Df(a) + \beta Dg(a). \end{aligned}$$

More impressive is the proof of the chain rule.

Theorem 4. *If f is differentiable at a and g is differentiable at $f(a)$ then $g \circ f$ is differentiable at a and*

$$D(g \circ f)(a) = Dg(f(a))Df(a).$$

Proof: Let $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ be differentiable at $a \in \mathbb{R}^n$ and let $g: \mathbb{R}^m \rightarrow \mathbb{R}^n$ be differentiable at $f(a) \in \mathbb{R}^m$. Then

$$\begin{aligned} g(f(x)) - g(f(a)) &= \psi(f(x))(f(x) - f(a)) \\ &= \psi(f(x))\phi(x)(x - a), \end{aligned}$$

where ϕ is a slope function for f at a and ψ is a slope function for g at $f(a)$.

Since ϕ and f are continuous² at a and ψ is continuous at $f(a)$ we have that $g \circ f$ is differentiable at a . Even more

$$D(g \circ f)(a) = \psi(f(a))\phi(a) = Dg(f(a))Df(a).$$

To finish this paragraph let us prove the critical point theorem.

Theorem 5. *Let f be a real-valued function defined on an open set $U \subset \mathbb{R}^n$. If f is differentiable at $x_0 \in U$ and $f(x_0)$ is a extreme value then $Df(x_0) = 0$.*

Proof: Suppose that $x_0 \in U$ is such that $f(x_0) \leq f(x)$ for all $x \in U$ and that f is differentiable at x_0 . Then

$$0 \leq f(x) - f(x_0) = \phi(x)(x - x_0) \quad (4)$$

for all $x \in U$, where ϕ is a slope function for f at x_0 . Let h be a fixed point in \mathbb{R}^n , and ε small enough so that $x_0 + th \in U$ for all $t \in (-\varepsilon, \varepsilon)$. By (4) we have that

$$t\phi(x_0 + th)h \geq 0$$

for all $t \in (-\varepsilon, \varepsilon)$. Hence

$$\phi(x_0 + th)h \geq 0, \quad t > 0,$$

and

$$\phi(x_0 + th)h \leq 0, \quad t < 0,$$

Since ϕ is continuous at x_0 we get $\phi(x_0)h = 0$. But h was chosen arbitrarily, then $\phi(x_0) = 0$, i.e., $Df(x_0) = 0$.

Observe that formally there is a little difference between the proofs of theorems 3, 4 and 5 and those in [Kuh]. Compare them also with the proofs of the same in [Spi].

To finish this paragraph we invite the reader to compute with this new formulation the derivative of a constant function, the derivative of a linear function and to prove the Leibniz rule.

5. THE INVERSE FUNCTION THEOREM. We do not pretend our proof of the inverse function theorem to be easy, but we believe ours is more direct than most of the proofs of the same we have seen (see for example [Spi], [Apo]), and that the reader can follow it without getting lost in too many details.

We start by extending the concept of differentiability at a point to differentiability in an open set. Let U be an open set in \mathbb{R}^n and $f: U \rightarrow \mathbb{R}^m$. We say that f is differentiable in U if for each $y \in U$ there is a slope function ϕ_y for f at y . In such a case we write $\phi(x, y) = \phi_y(x)$. Observe that $\phi(x, x) = Df(x)$. In the following we denote $Df(x)$ by λ_x .

²Continuity of f follows immediately from (2).

Let us define now continuous differentiability:

f is continuously differentiable in U if there exists a continuous function $\psi: U \times U \rightarrow M_{m \times n}$ such that

$$(5)$$

$$f(x) - f(y) = \psi(x, y)(x - y).$$

Observe that our definition is not the same as the conventional one:

f is continuously differentiable in U if f is differentiable in U and Df is continuous in U .

$$(6)$$

One sees clearly that (5) implies (6). Let us show that (6) implies (5). Suppose that f satisfies (6) and define $\psi: U \times U \rightarrow M_{m \times n}$ by

$$\psi(x, y) = \begin{cases} \frac{1}{\|x - y\|^2} \{(f(x) - f(y) - \lambda_x(x - y)) \otimes (x - y)\} + \lambda_x, & x \neq y \\ \lambda_x, & x = y. \end{cases}$$

It is clear that $\psi(x, y)(x - y) = f(x) - f(y)$ and that $\psi|_{U \times U - \Delta}$ and $\psi|_{\Delta}$ are continuous, where Δ is the diagonal of $U \times U$. What it is left to prove is that ψ is continuous on Δ . If $a \in U$,

$$\begin{aligned} \|\psi(x, y) - \psi(a, a)\| &\leq 2\|\lambda_x - \lambda_a\| + \frac{\|f(x) - f(y) - \lambda_a(x - y)\|}{\|x - y\|} \\ &\leq 2\|\lambda_x - \lambda_a\| + \|\lambda_{\xi} - \lambda_a\| \end{aligned}$$

where ξ is in the segment $[x, y]$. The last inequality is obtained by the mean value theorem. We get then, from the continuity of λ , the continuity of ψ . Therefore f satisfies (5).

With definition (5) we can now prove the inverse function theorem.

Theorem 6. *Let $f: U \rightarrow \mathbb{R}^n$ be defined in an open set $U \subset \mathbb{R}^n$. If f is continuously differentiable in U and $\det \psi(a, a) \neq 0$ for some point $a \in U$, then there exist neighborhoods V of a and W of $f(a)$ such that $f^{-1}: W \rightarrow V$ exists, is continuously differentiable and*

$$(Df^{-1})(f(a)) = [Df(a)]^{-1}.$$

Proof: We give the proof in three steps : (A) f is injective in some neighborhood of a , (B) f is onto some neighborhood of $f(a)$, and (C) f^{-1} is continuously differentiable in some neighborhood of $f(a)$.

Let ψ be as in (5) and assume that $\det \psi(a, a) \neq 0$, $a \in U$.

A) By continuity of ψ there is a neighborhood V of a such that $\det \psi(x, y) \neq 0$, $\psi(x, y)$ is bounded and $[\psi(x, y)]^{-1}$ is bounded for all $x, y \in V$. By (5) we conclude immediately that f is injective in V .

B)³ Let B_1 , be an open ball centered at a and contained in V . By continuity of f , $f(\partial B_1)$ ⁴ is compact. By A , $f(a) \notin f(\partial B_1)$. Let d be the distance from $f(a)$ to

³This step is taken from [Spi].

⁴ ∂B_1 denotes the border of B_1 .

$f(\partial B_1)$ and let B_2 be the open ball centered at $f(a)$ and radius $d/2$. Let us fix $y \in B_2$ and define the function g on \bar{B}_1^5 by

$$g(x) = \|y - f(x)\|^2.$$

Then g reaches its minimum at some point $x_0 \in \bar{B}_1$ since it is continuous and \bar{B}_1 is compact. By the definition of B_2 , x_0 is not in ∂B_1 . Now, g is differentiable at x_0 since

$$\begin{aligned} g(x) - g(x_0) &= \|y - f(x)\|^2 - \|y - f(x_0)\|^2 \\ &= (2y - f(x) - f(x_0))\psi(x, x_0)(x - x_0) \end{aligned}$$

and $\phi(x) = (2y - f(x) - f(x_0))\psi(x, x_0)$ is continuous at x_0 . By Theorem 5

$$0 = \phi(x_0) = 2(y - f(x_0))\psi(x_0, x_0)$$

and therefore $y = f(x_0)$ (see A). Thus f maps B_1 onto B_2 .

C) Restrict f to $W = f^{-1}(B_2) \cap B_1$. By A and B, $f: W \rightarrow B_2$ is bijective and then we can define $f^{-1}: B_2 \rightarrow W$. By (5), for $z, w \in B_2$ we have

$$f^{-1}(z) - f^{-1}(w) = [\psi(f^{-1}(z), f^{-1}(w))]^{-1}(z - w). \quad (7)$$

Now, since $[\psi(f^{-1}(z), f^{-1}(w))]^{-1}$ is bounded (see A), f^{-1} is continuous and therefore continuously differentiable in B_2 by (7). Also by (7)

$$(Df^{-1})(f(a)) = [Df(a)]^{-1}.$$

We invite the reader to compare Spivak's proof of this theorem with our proof. One can see that the proof of step (B) is the same in both, but the proofs of parts (A) and (C), especially part (C), clearly show the advantages of Carathéodory's formulation.

CONCLUSIONS. From a pedagogical point of view Carathéodory's characterization is interesting. In \mathbb{R} , it enhances the geometrical sense of the derivative as the continuous approximation to the tangent line by means of secant lines to the graph of a function, showing that continuity is an essential element for differentiability. This formulation allows us to prove with economy and elegance, the elementary facts of differentiability, not only of real functions of real variable, but also of functions from \mathbb{R}^n to \mathbb{R}^m where the arguments are the same ones used for functions in one variable.

ACKNOWLEDGMENTS. We want to thank Carlos Rodriguez and Jürgen Tischer for their interest and helpful suggestions. We also thank Luis Recalde who pointed out the reference [Apo] where Apostol uses Carathéodory's formulation for real functions of one real variable.

REFERENCES

-
- [Aco,Del] E. Acosta, C. Delgado, *La Derivada de Carathéodory en Espacios Vectoriales Topológicos*, Preprint.
[Apo] Tom M. Apostol, *Análisis Matemático*, Editorial Reverté, Barcelona, (1986).
[Car] Constantin Carathéodory, *Theory of Functions of a Complex Variable*, Vol. I, Chelsea, New York, (1954).

⁵ \bar{B}_1 denotes the closure of B_1 .

- [Fre] M. Fréchet, *La notion de différentielle dans l'analyse générale*, C. R. Acad. Sci. (Paris), 180 (1925), 806–909.
- [Kuh] Stephen Kuhn, *The Derivative a la Carathéodory*, *American Mathematical Monthly*, Vol. 98, No. 1, January, (1991).

Departamento de Matemáticas
Universidad del Valle
Apártado Aéreo 2188
Santiago de Cali
COLOMBIA

$2^{858433} - 1$ is Prime

David Slowinski and Paul Gage have verified yet another Mersenne prime. Using a Cray C916 at the Cray Research computing facility, they verified that $2^{858433} - 1$ is prime. The run time on 16 processors was about 30 minutes.

To verify that a random integer of this size is prime would be well beyond the range of present computers (as well as those for years to come). Mersenne primes, however, are relatively easy to test using the Lucas-Lehmer test. To test $N = 2^p - 1$, start with 4 and repeatedly square and subtract 2. Then N is prime precisely when, after $p - 2$ repetitions, the result is 0 modulo N .

Sounds easy. The problem is squaring (and reducing) numbers with thousands of digits. Performing multiprecision multiplication quickly requires some sophistication, and Slowinski and Gage use an algorithm of Schonhage and Strassen, implemented using the Fast Fourier Transform.

Why do they continue the search? Running this particular set of programs seems to be a good test of reliability for both hardware and software on new machines. In addition, of course, there is a fascination with finding the largest known prime—at least, largest for the moment.

Slowinski and Gage comment that they have not checked all exponents *below* 858,433. There may be other Mersenne primes to be discovered. But chances are that people who have the time and inclination (and a Cray) will forge on to higher values. After all, there's nothing remarkable about finding the *second* largest known prime number.

Odd Magic Powers

A. C. Thompson

INTRODUCTION. By a magic square we understand a square array of numbers (a square matrix) whose row sums, column sums and two diagonal sums are all equal. The magic square

$$\begin{bmatrix} 10 & 5 & 11 & 8 \\ 3 & 16 & 2 & 13 \\ 6 & 9 & 7 & 12 \\ 15 & 4 & 14 & 1 \end{bmatrix}$$

(which is obtained from the famous example in Dürer's *Melancholia* by interchanging the first two rows and the first two columns) has the additional property that all of its 8 diagonals (e.g. 5, 2, 12, 15; 11, 13, 6, 4; 10, 13, 7, 4) also have the same sum. A magic square with this extra property is called a *pandiagonal* magic square. Conventionally, additional restrictions are placed on the entries: (a) that they be (non-negative) integers; and often, (b) that they be the consecutive integers from 1 to n^2 (when the array is $n \times n$). We shall not require either of these restrictions and the numbers may be rational, real or complex.

The purpose of this note is to consider the multiplicative properties of sets of magic squares (using the normal matrix multiplication). As van den Essen [3] has shown, some of these turn out to be quite surprising. We extend his results by showing that the product of any odd number of 3×3 magic squares is magic, that if such a square is invertible, then the inverse is magic and that these results extend to 4×4 and 5×5 pandiagonal magic squares. Because our method of proof is somewhat different from that of [3] and for completeness, we present the whole argument. As C. Small [2] points out, results of this type provide good examples for a Linear Algebra class.

The first question of a multiplicative type is to ask whether a magic square is necessarily invertible. Clearly the answer is no—take the trivial case with all entries equal (even 0!) but a less trivial one is the example above (the vector $(3, -1, -3, 1)^T$ is in the null space).

Our notation and terminology is standard except that we shall use *first* and *second* diagonal to mean the main diagonal and the 'other' one respectively. Also, we shall call a matrix with constant row sums an *affine* matrix and an affine matrix with constant column sums will be called *doubly affine*. Since we often need various permutation matrices, these will be defined by listing the columns as elements of the usual basis $\{e_1, e_2, \dots, e_n\}$ of \mathbb{R}^n . For example:

$$P = [e_3, e_2, e_1] = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Proposition 1. *Both the set of $n \times n$ affine matrices and the set of $n \times n$ doubly affine matrices are algebras.*

Proof: It is clear that both sets are closed under addition and scalar multiplication. Moreover, a matrix is affine if and only if the constant vector $(1, 1, \dots, 1)^T$ is an eigenvector with eigenvalue equal to the row sum r . It follows that the set of affine matrices is closed under multiplication. The same fact for doubly affine matrices follows by considering transposes.

Thus the crucial part of our problem is to consider the behaviour of diagonal sums under multiplication.

1. THE 3×3 CASE. This has been discussed (in an entirely numerical way) by J. M. H. Peters in [1].

Proposition 2. *If A is a 3×3 invertible magic square, then A^{-1} is magic.*

Proof: Let r be the row sum of A . Then r is an eigenvalue of A (and of A^T) and $\text{tr } A = r$. Hence, if λ_1 and λ_2 are the remaining eigenvalues we have $\lambda_1 + \lambda_2 = 0$. Since A is invertible $\lambda_1 = -\lambda_2 \neq 0$. Hence, $\lambda_1^{-1} = -\lambda_2^{-1}$ and $\text{tr } A^{-1} = r^{-1} + \lambda_1^{-1} + \lambda_2^{-1} = r^{-1} = \text{row sum of } A^{-1}$.

For the second diagonal we consider $B = PA$ where $P = [e_3, e_2, e_1]$. Clearly B is also magic with the same row sum r . Therefore, the same argument shows that $\text{tr}(B^{-1}) = r^{-1}$. However, $B^{-1} = A^{-1}P^{-1} = A^{-1}P$ and hence $\text{tr}(B^{-1})$ is the second diagonal sum of A^{-1} .

The 3×3 magic squares form a 3-dimensional space with basis $\{E_1, E_2, E_3\}$ where

$$E_1 = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

The larger algebra \mathcal{A} of 3×3 doubly affine matrices is 5-dimensional (now the two diagonal sums can vary) and has a basis consisting of $E_1, E_2, E_3, I = [e_1, e_2, e_3]$ and $P = [e_3, e_2, e_1]$.

Theorem 1. *The product of an odd number of 3×3 magic squares is again a magic square.*

Proof: First look at products of pairs from the basis $\{E_1, E_2, E_3\}$. We get $E_1E_3 = E_3E_1 = E_2E_3 = E_3E_2 = 0$; $E_1E_2 = -E_2E_1 = 3P - E_3$; $E_1^2 = -E_2^2 = E_3 - 3I$ and $E_3^2 = 3E_3$. Hence the product of any two 3×3 magic squares is in the subspace of \mathcal{A} spanned by I, P , and E_3 . But this subspace is a subalgebra. Hence the product of any even number of magic squares is in this subalgebra. Finally, since the product of any magic square with each of E_3, I , and P is magic, we have the result.

Corollary. *If A is a 3×3 magic square then A^k is magic for every odd positive integer; and for every odd integer if A is invertible.*

2. THE 4×4 CASE. As in the previous case the fact that the trace is equal to one of the eigenvalues means that the sum of the other three is 0. This can happen in three ways.

- (i) The remaining eigenvalues are $0, \pm \lambda$. It is an interesting exercise to show that this is so for all 4×4 pandiagonal magic squares. Thus none of these

matrices is invertible. It is, however, true that a product of an odd number of 4×4 pandiagonal magic squares is pandiagonal magic. The proof of this is similar to, but less natural than, the 5×5 case. For reasons of space we omit it.

- (ii) The remaining eigenvalues are $\lambda, \omega\lambda, \omega^2\lambda$ (where ω is a complex cube root of unity). An example of such a magic square is

$$A = [e_1, e_4, e_2, e_3]$$

for which $A^2 = A^{-1}$ and is magic. So it is possible for a product of an even number of magic squares to be magic.

- (iii) The remaining eigenvalues are not symmetric with respect to the origin. An example is

$$\begin{bmatrix} 10 & 2 & 14 & 18 \\ 17 & 15 & 1 & 11 \\ 13 & 21 & 7 & 3 \\ 4 & 6 & 22 & 12 \end{bmatrix}.$$

3. THE 5×5 CASE. The previous cases show that magic squares with nice multiplicative properties have their eigenvalues (other than the row sum r) symmetrically placed with respect to the origin. To achieve this when n is odd we need the characteristic polynomial to factor into $(x - r)p(x)$ where $p(x)$ is even. It appears that, as n increases, increasingly many extra conditions are needed to guarantee such a factorization.

The invertible magic square

$$\begin{bmatrix} 23 & 1 & 2 & 20 & 19 \\ 22 & 16 & 9 & 14 & 4 \\ 5 & 11 & 13 & 15 & 21 \\ 8 & 12 & 17 & 10 & 18 \\ 7 & 25 & 24 & 6 & 3 \end{bmatrix}$$

is an example of one whose inverse is not magic. It is also not pandiagonal. We will see below that the extra pandiagonal condition is the “right” one to impose on 5×5 magic squares to recover the results of the 3×3 case.

Another linear algebra exercise is to show that the 5×5 pandiagonal magic squares form a 9-dimensional vector space. To describe a basis for this space we consider the following 10 matrices:

$$P_1 = [e_1, e_4, e_2, e_5, e_3], P_2 = [e_4, e_2, e_5, e_3, e_1], \dots, P_5 = [e_3, e_1, e_4, e_2, e_5],$$

$$Q_1 = [e_3, e_5, e_2, e_4, e_1], Q_2 = [e_5, e_2, e_4, e_1, e_3], \dots, Q_5 = [e_1, e_3, e_5, e_2, e_4];$$

where, in each line, the columns are permuted cyclically. Each of these matrices is pandiagonal and, because $\sum_{i=1}^5 P_i = \sum_{i=1}^5 Q_i$ and this is the only non-trivial linear condition they satisfy, their span is 9-dimensional. Thus every pandiagonal 5×5 magic square is of the form

$$A = \sum_{i=1}^5 \alpha_i P_i + \sum_{i=1}^5 \beta_i Q_i.$$

Theorem 2. *The product of an odd number of 5×5 pandiagonal magic squares is again a pandiagonal magic square.*

Proof: Consider the algebra \mathcal{B} spanned by R and S where

$$R = [e_5, e_4, e_3, e_2, e_1] \quad \text{and} \quad S = [e_2, e_3, e_4, e_5, e_1].$$

This algebra is also 9-dimensional being the linear span of

$$S, S^2, S^3, S^4, S^5 = I; RS, RS^2, RS^3, RS^4, RS^5 = R;$$

(again $\sum_{i=1}^5 S^i = \sum_{i=1}^5 RS^i$). These 10 matrices themselves form a group isomorphic to the symmetry group of a regular pentagon).

Now it is an elementary matter to check that, for $i, j = 1, 2, \dots, 5$ we have both $P_i Q_j = S^k$ and $Q_i P_j = S^{k'}$ for some value of k and k' depending on i and j . Similarly, $P_i P_j = RS^m$ and $Q_i Q_j = RS^{m'}$. Thus if A_1 and A_2 are two pandiagonal magic squares then $A_1 A_2$ is in \mathcal{B} . Hence every product of an even number of such squares is in \mathcal{B} . Multiplying a matrix on the right by a power of S permutes the columns cyclically and a similar multiplication by R reverses the order of the columns. Clearly the set $\{P_i, Q_i; i = 1, 2, \dots, 5\}$ is invariant under these actions. Therefore, if A is a 5×5 pandiagonal magic square so is AB for every B in \mathcal{B} ; from which the result follows.

Corollary 1. *If A is a 5×5 pandiagonal magic square, so is A^k for every odd positive integer k .*

Corollary 2. *If A is a 5×5 pandiagonal magic square and if $\lambda_1, \lambda_2, \lambda_3$ and λ_4 denotes its eigenvalues other than the row sum r , then (suitably numbered) $\lambda_1 = -\lambda_2$ and $\lambda_3 = -\lambda_4$.*

Proof: It follows immediately from Corollary 1 that $\text{tr } A^k = r^k$ and hence that $\sum_{i=1}^4 \lambda_i^k = 0$ for every odd positive integer k . If we take $k = 1$, and $k = 3$, and use elementary algebra we get the result.

Corollary 3. *If A is an invertible 5×5 pandiagonal magic square so is A^{-1} .*

Proof: The proof follows from Corollary 2 exactly as in the proof of Proposition 2, but using the permutation matrices R and S in place of P .

Note. This is a revised version of a paper originally submitted in October 1988.

REFERENCES

1. J. M. H. Peters, Inverses and cubes, Math. Gazette 65 (1981) 253–4.
2. C. Small, Magic Squares over fields, Amer. Math. Monthly 95 (1988) 621–625.
3. A. van den Essen, Magic Squares and Linear Algebra, Amer. Math. Monthly (1990) 60–62.

*Department of Mathematics
Dalhousie University
Halifax, Nova Scotia
B3H 3J5 Canada
tony@cs.dal.ca*

**Answer to Who Was the Author
(p. 317)
George Abraham Miller**

Mathematicians, Including Undergraduates, Look at Soap Bubbles

Frank Morgan

Why are soap bubbles so beautifully round? A soap bubble tries to minimize surface energy or area, and the round sphere has the least surface area for the fixed volume of air trapped inside. See FIGURE 1. A soap bubble very quickly succeeds in finding this mathematically optimal shape. Thus the underlying principle is *area minimization*.

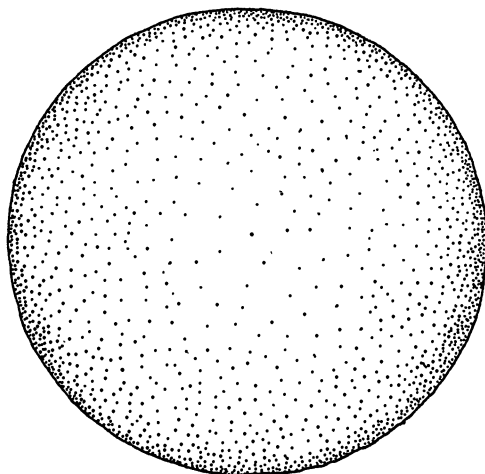


Figure 1. A spherical soap bubble has found the least-area way to enclose a given volume of air. Jim Bredt [M4].

Similarly bubble clusters try to minimize the total surface area enclosing and separating several volumes. Whether the number of enclosed volumes is one or one thousand, it is the same principle of area minimization at work. See FIGURES 2, 3. This principle alone has sufficed to produce computer simulations of bubble clusters, as in the frame in FIGURE 4 from the video “Computing soap films and crystals” by the Minimal Surface Team at the Geometry Center (formerly the Minnesota Geometry Supercomputer Project).

Do soap bubble clusters always find the absolute least-area shape? Not always: FIGURE 5 illustrates two clusters enclosing and separating the same five volumes. In the first, the tiny fifth volume is comfortably nestled deep in the crevice between the largest bubbles. In the second, the tiny fifth volume less comfortably sits between the medium size bubbles. The first cluster has less surface area, although I do not know for sure that there is not a third possibility of still less area.

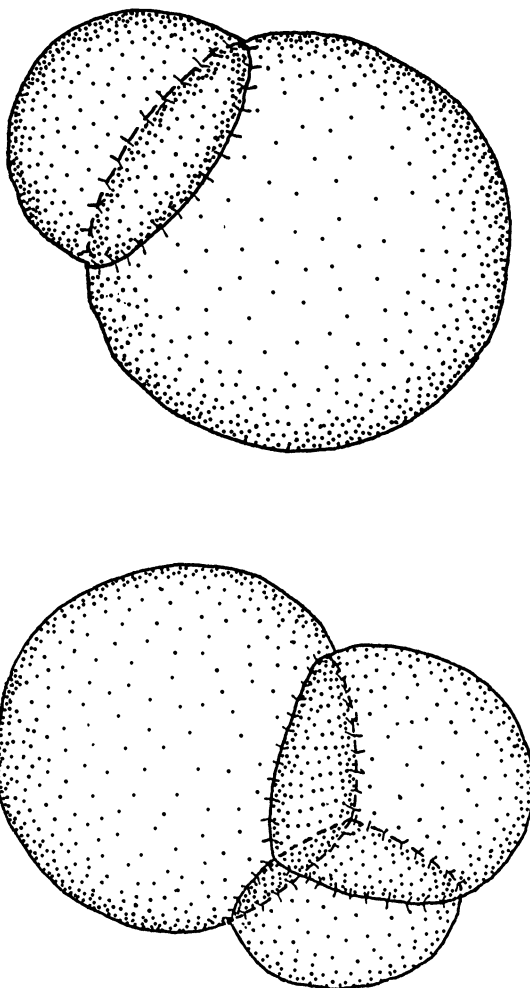


Figure 2a and 2b. Double and triple bubbles seek the least-area way to enclose and separate two or three given volumes of air.

As a matter of fact, it is an open question whether the standard triple bubble of FIGURE 2 is the least-area way to enclose three given volumes. Even for the double bubble, the proof seems incomplete, as realized over the course of a recent undergraduate thesis [F1] by Joel Foisy, Williams '91.

In fact, for area-minimizing clusters, it is an open question whether each separate region is connected, or whether it might conceivably help to subdivide the regions of prescribed volume, with perhaps half the volume nestled in one crevice here, and the other half in another crevice there.

Similarly, it is an open question whether an area-minimizing cluster may incidentally trap inside “empty chambers,” which do not contribute to the prescribed volumes. FIGURE 6 shows a 12-bubble with a dodecahedral-like empty chamber on the inside, obtained by Tyler Jarvis of Princeton University using a program of Ken Brakke [B] of the Geometry Center. The computation postulated the empty chamber; without such a restriction, empty chambers probably never occur.

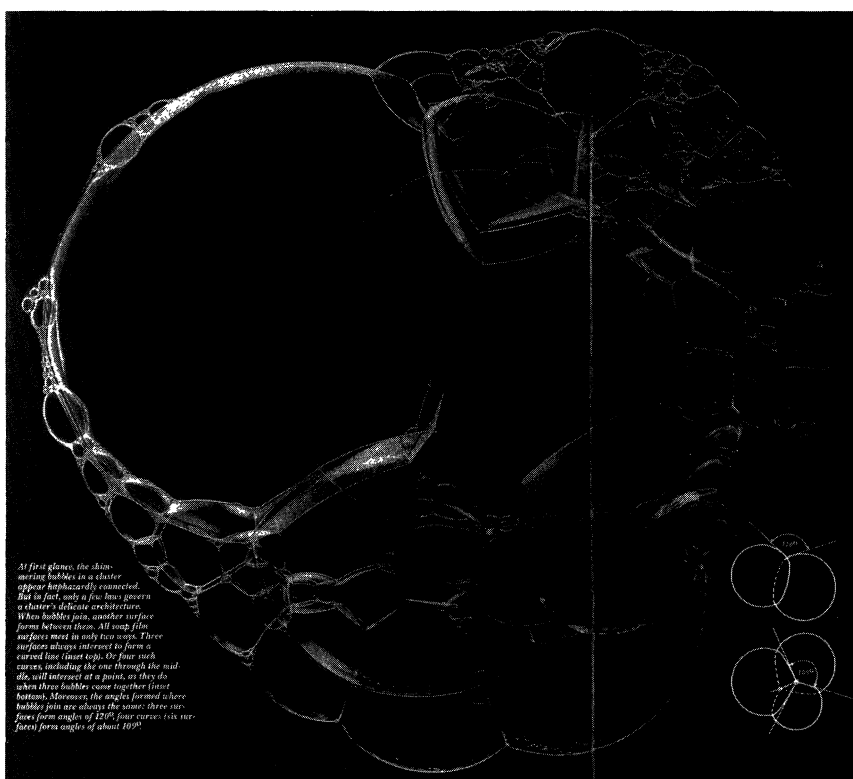


Figure 3. Configurations of thousands of bubbles are governed by the same principle of area minimization. Photograph from Science '84 [8] courtesy of Gordon Graham/Prism.

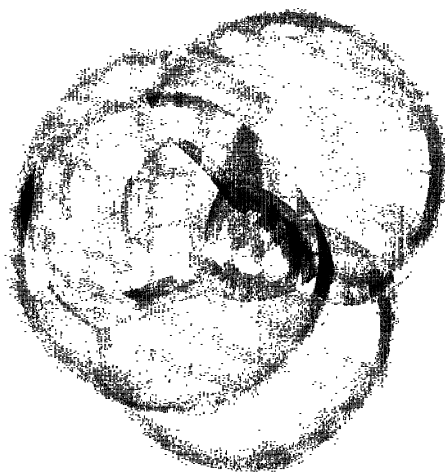


Figure 4. A computer simulation of a bubble cluster. “Computing soap films and crystals” video by the Minimal Surface Team, The Geometry Center. Photo provided by John Sullivan.

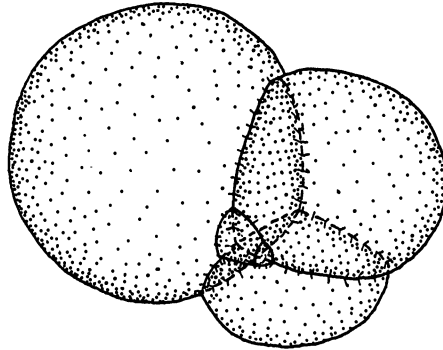
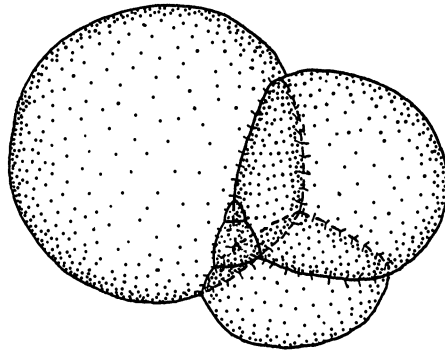


Figure 5. Soap bubble clusters are sometimes only relative minima for area. These two clusters enclose and separate the same five volumes, but the first has less surface area than the second.

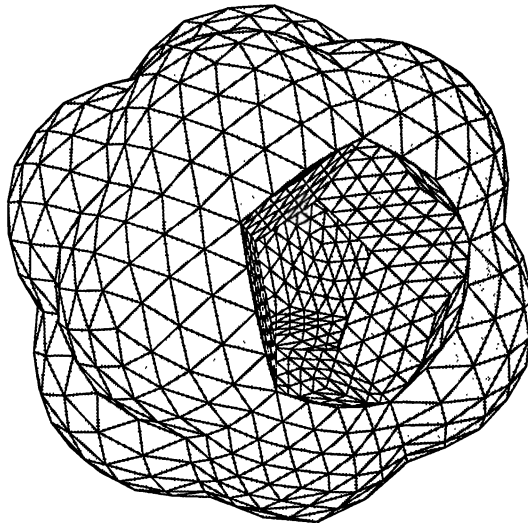


Figure 6. It is an open question whether area-minimizing clusters may have empty chambers, such as the dodecahedral-like chamber at the center of this 12-bubble. Tyler Jarvis, Princeton University.

You must be wondering what is known! It is known that soap bubble clusters consist of constant-mean-curvature surfaces (such as pieces of spheres) meeting in threes at 120° angles along seams, which in turn meet in fours at about 109° angles at points. Four such seams and two such points are visible already in the triple bubble of FIGURE 2. Nothing more complicated ever happens, even in such complicated clusters as that of FIGURE 3.

These laws were observed and recorded by the Belgian physicist J. A. F. Plateau [P] over a century ago, but it was over 100 years until a complete explanation was proved by Jean Taylor, now a professor at Rutgers University. Her demonstration required no physics or chemistry, just a single mathematical hypothesis: *area-minimization*. Many pages of complicated mathematics later came the conclusion: Plateau's laws, 120° angles, 109° angles, and all.

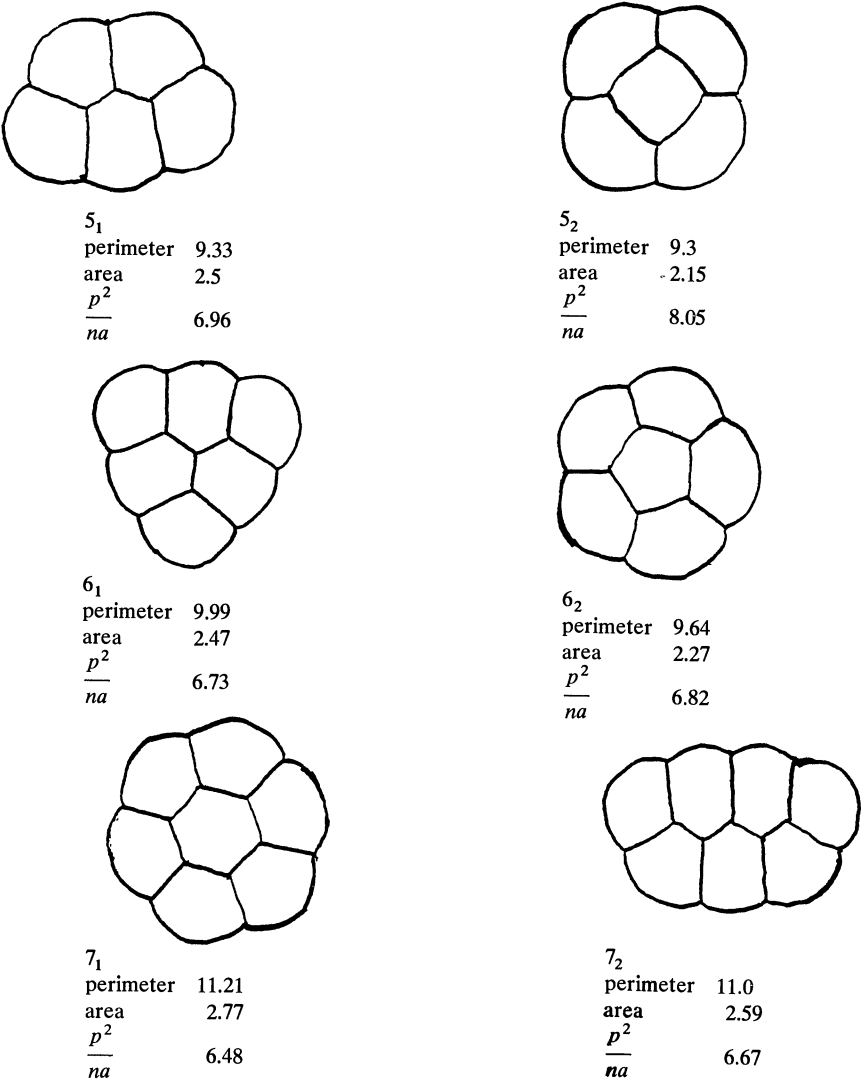


Figure 7a. Comparison of candidates for the least-perimeter way to enclose and separate $n = 5, 6, 7$ unit planar areas. Note that only sometimes are the most symmetric candidates the winners.

Taylor's work used *geometric measure theory*, a relatively new kind of geometry that allows singularities such as the seams in soap bubble clusters (cf. [M2]). Classically, geometry studied mainly smooth surfaces.

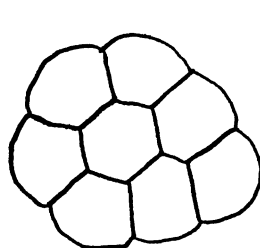
Although Taylor's mathematical proof [T] is hard to read, a beautiful expository account [AT] appeared in *Scientific American* in July, 1976.

Joel Foisy, whose undergraduate thesis work on the double bubble I mentioned earlier, got started the previous summer with an undergraduate research Geometry Group at Williams studying the

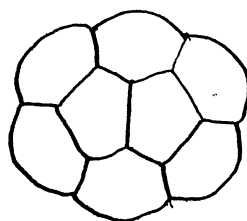
Planar soap bubble problem. Find the least-perimeter way to enclose and separate n given areas (say unit areas) in the plane.

FIGURES 7ab show the results of their experiments with soap bubbles between plexiglass plates [A]. Since it is hard to keep all the areas exactly 1, it is better to minimize the normalized quantity p^2/na , where p is the total perimeter (counting all inside and outside walls once) and a is the total area. Could you have guessed the winners? I find it hard to see any pattern.

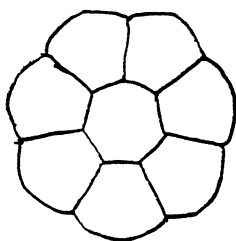
There was also one notable theoretical result:



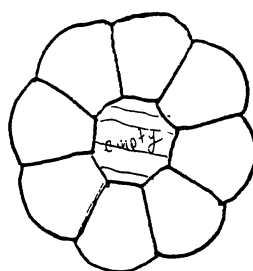
8_1	
perimeter	12.01
area	2.85
$\frac{p^2}{na}$	6.33



8_2	
perimeter	12.45
area	2.97
$\frac{p^2}{na}$	6.52



8_3	
perimeter	12.57
area	3.03
$\frac{p^2}{na}$	6.52



8_4	
perimeter	14.69
area	3.29
$\frac{p^2}{na}$	8.2

Figure 7b. Comparison of four candidates for the least-perimeter way to enclose and separate eight unit areas. For the fourth candidate, the eight enclosed areas incidentally surround and enclose an extraneous "empty chamber."

81 Stanton Road
 Brookline, MA 02146
 January 2, 1983

Professor Frank Morgan
 Rice University
 Department of Mathematics
 Houston, Texas 77001

Dear Professor Morgan:

Thank you again for sending me the 18.01-18.02 MIT Calculus materials. I recently took my first 18.02 test and made no errors on it.

I believe that I have discovered a solution to the soap film problem which was the topic of your speech at the 1982 Massachusetts State Science Fair, namely, *Find a simple, smooth, closed curve which can bound infinitely many minimal surfaces.*

Start with the shape in Figure 1, a shape that has two minimal surfaces,

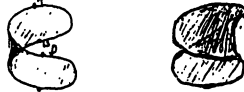


Figure 1.

and elongate it at points A, B, C, and D to create the shape in Figure 2,

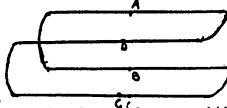


Figure 2.

and coat both of the two minimal surfaces with bubble suds. Introduce crosspieces AB, BC, CD, and DA and coat the resultant square with bubble suds too. By popping the proper bubble films you will get the configuration portrayed in Figure 3.



Figure 3.

Remove the crosspieces with care so as to not pop any more of the bubble films, and the shape will settle into the saddle of Figure 4.



Figure 4.

This saddle can be shifted along the entire length of the shape, creating infinitely many minimal surfaces.

I built a model of Figure 4 and found that the saddle could be slid along the length of the figure by tilting one end towards the floor. (The weight of the bubble suds made it slide - I wonder what would happen in a weightless condition.)

I have been unable to form a minimal surface on a mobius strip - I think that this happens because there is only one edge or surface for the bubble suds to adhere to. (Am I right?)

Sincerely yours,

Mark Kantrowitz

Mark Kantrowitz

Figure 8. As a high-school sophomore, Mark Kantrowitz addresses an open question on minimal surfaces.

Theorem ([A], [F1], [F2]). *The standard double bubble is uniquely perimeter minimizing.*

The hard part is showing that both areas are connected and that there are no empty chambers. For more bubbles or in space, such questions remain open.

Undergraduate research. Undergraduates do mathematics, prove theorems, write papers for publication, and give talks at mathematics meetings. In the Williams College SMALL undergraduate research project, each student belongs to two of seven research groups, each with a student leader and a faculty advisor.

The student groups work largely independently, although the faculty advisor usually provides the problem and some guidance. The students generally are quite busy and do not like to be interrupted by the faculty! The less experienced students learn from the others, and the most advanced like to bounce ideas off the rest.

One woman, who had no intentions of attending graduate school, changed her mind during the summer project and is now a graduate student in computer science at Wisconsin. A nonmajor, who used to consider mathematics just an avocation, is now a graduate student in Mathematics at UC Berkeley. A Hispanic student, who never seemed to settle down in his coursework, succeeded outstandingly in the SMALL project his freshman year. In his junior year he wrote an Honors thesis, which has been accepted for publication in the *Pacific Journal of Mathematics*. He now has two jobs, teaching mathematics in his old high school and at a community college, including one course taught in Spanish.

The name SMALL, in case you were wondering, is an acronym for the faculty on the original proposal.

It is probably good to start research early on. One time I got a letter from Mark Kantrowitz, a high-school sophomore, that announced, “I believe that I have discovered a solution to the [open] soap film problem . . . : *Find a simple, smooth, closed curve which can bound infinitely many minimal surfaces.*” See FIGURE 8. Although Mark’s stated proposal used gravity, the basic idea was one I had used in at least one published example. Mark’s continuing correspondence amazed me. My collaborators and I barely managed to keep ahead of him. The problem is still open.

REFERENCES

-
- [A] Manuel Alfaro, Jeffrey Brock, Joel Foisy, Nickelous Hodges, and Jason Zimba, Compound soap bubbles in the plane, SMALL Geometry Group report, Williams College, 1990.
 - [AT] F. J. Almgren, Jr., and Jean E. Taylor, The geometry of soap films and soap bubbles, *Scientific American*, July, 1976, pp. 82–93.
 - [B] Kenneth A. Brakke, The surface evolver, *Experimental Math.* 1 (1992), 141–165.
 - [CR] R. Courant and H. Robbins, *What is Mathematics?*, Oxford Univ. Press, 1941.
 - [F1] Joel Foisy, Soap bubble clusters in \mathbf{R}^2 and \mathbf{R}^3 , Honors thesis, Williams College, 1991.
 - [F2] Joel Foisy, Manuel Alfaro, Jeffrey Brock, Nickelous Hodges, and Jason Zimba, The standard double soap bubble in \mathbf{R}^2 uniquely minimizes perimeter, *Pacific J. Math.*, 159 (1993), 47–59.
 - [M1] Frank Morgan, *Compound soap bubbles, shortest networks, and minimal surfaces*, AMS video, 1992.
 - [M2] Frank Morgan, *Geometric Measure Theory: a Beginner’s Guide*, Academic Press, 1988.
 - [M3] Frank Morgan, Minimal surfaces, crystals, shortest networks, and undergraduate research, *Math. Intel.*, Vol. 14, Summer, 1992, 37–44.
 - [M4] Frank Morgan, Soap bubbles and soap films, in Joseph Malkevitch and Donald McCarthy, ed., *Mathematical Vistas: New and Recent Publications in Mathematics from the New York Academy of Sciences*, Vol. 607, 1990.

- [P] J. A. F. Plateau, *Statique Experimentale et Theorique des Liquides Soumis aux Seules Forces Moleculaires*, Paris, Gauthier-Villars, 1873.
- [S] Bruce Schecter, Bubbles that bend the mind, *Science* 84, March, 1984.
- [T] Jean E. Taylor, The structure of singularities in soap-bubble-like and soap-film-like minimal surfaces, *Ann. of Math.* 103 (1976), 489–539.

This article is based on an AMS-MAA address in San Francisco, 1991, available on video [M1]. The second half appears as [M3]. The work was partially supported by the National Science Foundation and the Institute for Advanced Study.

Added in proof. There has been progress on the planar triple bubble by the 1992 Williams NSF SMALL undergraduate research Geometry Group, Chris Cox, group leader, Lisa Harrison, Michael Hutchings, Susan Kim, Janette Light, Andrew Mauer, Meg Tilton, “The shortest enclosure of three connected areas in \mathbf{R}^2 ” (preprint), and on the double bubble in space by Michael Hutchings. See the new chapter on “Soap bubble clusters,” in the second edition of [M2], to appear this year.

Department of Mathematics
Williams College
Williamstown, MA 01267
Frank.Morgan@williams.edu

“The advantage is that mathematics is a field in which one’s blunders tend to show very clearly and can be corrected or erased with a stroke of the pencil. It is a field which has often been compared with chess, but differs from the latter in that it is only one’s best moments that count and not one’s worst. A single inattention may lose a chess game, whereas a single successful approach to a problem, among many which have been relegated to the wastebasket, will make a mathematician’s reputation.”

Excerpt from *Ex-Prodigy:
 My Childhood and Youth*
 by Norbert Wiener, p. 21

NOTES

Edited by: John Duncan

A Proof of Dilworth's Chain Decomposition Theorem

Fred Galvin

Let P be a finite partially ordered set: a *chain* (*antichain*) in P is a set of pairwise comparable (incomparable) elements; the *width* of P is the maximum cardinality of an antichain in P . According to a celebrated theorem of Dilworth [2], the width of P is also equal to the minimum number of chains needed to cover P . The wider combinatorial significance of Dilworth's theorem, especially as regards matching theory, is discussed by Bogart, Greene, and Kung [1], Mirsky [3], and Reichmeider [5]. Bogart, Greene, and Kung survey various proofs of Dilworth's theorem; the proofs of Perles [4] and Tverberg [6] are especially simple and elegant. The proof given here seems to me to be as simple as any. This proof is probably well-known folklore; still, as far as I know, it has never appeared in print.

Theorem. *A finite partially ordered set P is the union of m chains, where m is the width of P .*

Proof: We use induction on the cardinality of P . Let a be a maximal element of P , and let $P' = P \setminus \{a\}$ have width n . Then P' is the union of n disjoint chains C_1, \dots, C_n . We have to show that P either contains an $(n + 1)$ -element antichain, or else is the union of n chains. Now, every n -element antichain in P' consists of one element from each C_i . Let $a_i = \max\{x \in C_i: x \text{ belongs to some } n\text{-element antichain in } P'\}$. It is easy to see that $A = \{a_1, \dots, a_n\}$ is an antichain. If $A \cup \{a\}$ is an antichain, we are done. Otherwise, we have $a > a_i$ for some i . Then $K = \{a\} \cup \{x \in C_i: x \leq a_i\}$ is a chain, and there are no n -element antichains in $P \setminus K$, whence $P \setminus K$ is the union of $n - 1$ chains.

REFERENCES

1. Kenneth P. Bogart, Curtis Greene, and Joseph P. S. Kung, The impact of the chain decomposition theorem on classical combinatorics, in: Kenneth P. Bogart, Ralph Freese, and Joseph P. S. Kung, Eds., *The Dilworth Theorems*, Birkhäuser, Boston, 1990, 19–29.
2. R. P. Dilworth, A decomposition theorem for partially ordered sets, *Ann. of Math.* 51 (1950), 161–166.
3. L. Mirsky, *Transversal Theory*, Academic Press, New York, 1971.
4. Micha A. Perles, A proof of Dilworth's decomposition theorem for partially ordered sets, *Israel J. Math.* 1 (1963), 105–107.
5. Philip F. Reichmeider, *The Equivalence of Some Combinatorial Matching Theorems*, Polygonal Publishing House, Washington, New Jersey, 1984.

Department of Mathematics
University of Kansas
Lawrence, KS 66045-2142

On Intervals, Transitivity = Chaos.

Michel Vellekoop and Raoul Berglund

In an earlier article in the *American Mathematical Monthly* a redundancy was found in the definition of chaos by Devaney [1]:

Let V be a set. A continuous map $f: V \rightarrow V$ is said to be chaotic on V if

- (1) f is *topologically transitive*: for any pair of open non-empty sets $U, W \subset V$ there exists a $k > 0$ such that $f^k(U) \cap W \neq \emptyset$.
- (2) the *periodic points* of f are *dense* in V .
- (3) f has *sensitive dependence on initial conditions*: there exists a $\delta > 0$ such that, for any $x \in V$ and any neighbourhood N of x , there exists a $y \in N$ and an $n \geq 0$ such that $|f^n(x) - f^n(y)| > \delta$.

In [2], Banks et al. prove that (1) and (2) imply (3) in any metric space V , and in [3] Assaf IV and Gadbois show that for *general* maps this is the only redundancy: (1) and (3) do not imply (2), and (2) and (3) do not imply (1). But if we restrict our attention to maps on an interval a stronger result can be obtained:

Proposition. *Let I be a, not necessarily finite, interval and $f: I \rightarrow I$ a continuous and topologically transitive map. Then (1) the periodic points of f are dense in I and (2) f has sensitive dependence on initial conditions.*

The first result (1) can be found in [4] (Chapter IV.5, Lemma 41) but the proof uses a lot of other highly non-trivial results. Since Devaney's text is being used by so many students, we think that it is interesting to give a very short, intuitive proof of this proposition.

We will need the following lemma, which can be found in [4] (Chapter IV.1, Corollary 10) in a more general form:

Lemma. *Suppose that I is a, not necessarily finite, interval and $f: I \rightarrow I$ is a continuous map. If $J \subset I$ is an interval which contains no periodic points of f and $z, f^m(z)$ and $f^n(z) \in J$ with $0 < m < n$, then either $z < f^m(z) < f^n(z)$ or $z > f^m(z) > f^n(z)$.*

Proof of the lemma: Suppose we can find such a $z \in J$ with $z < f^m(z)$ and $f^m(z) > f^n(z)$. Define the function $g(x) = f^m(x)$. Then we know that $z < g(z)$ and this implies $z < g(z) < g^{k+1}(z)$ for all natural numbers $k \geq 1$ by induction. Because, if $g^{k+1}(z) < g(z)$ for a certain k then the function $g^k(x) - x$ has a

positive value in z and a negative value in $g(z)$ and this would mean, by the Intermediate Value Theorem, that there exists a point $c \in]z, g(z)[\subset J$ with $g^k(c) - c = 0$, giving a km -periodic point of f in J . Thus $z < g^k(z)$ for all positive k so in particular for $k = n - m > 0$, giving $z < f^{(n-m)m}(z)$. Since we assumed that $f^{n-m}(f^m(z)) < f^m(z)$ we could prove analogously, taking $g = f^{n-m}$ that $f^{(n-m)m}(f^m(z)) < f^m(z)$. But then we have that the function $f^{(n-m)m}(x) - x$ has a positive value in z and a negative value in $f^m(z)$, giving an $(n - m)m$ -periodic point in J and thus a contradiction. The other case can be proven analogously. \square

Proof of the Proposition: Suppose that f is continuous and topologically transitive. Because of the result in [2] we only need to prove that the periodic points are dense in I . Suppose that this is not the case, then there exists an interval $J \subset I$ containing no periodic points. Take an $x \in J$ which is not an endpoint of J , an open neighbourhood $N \subsetneq J$ of x and an open interval $E \subset J \setminus N$. Since f is topologically transitive on I there exists a natural number $m > 0$ with $f^m(N) \cap E \neq \emptyset$ and thus a $y \in J$ with $f^m(y) \in E \subset J$. Since J contains no periodic points we know that $y \neq f^m(y)$ and since f is continuous this implies that we can find a neighbourhood U of y with $f^m(U) \cap U = \emptyset$. Since U is an open set we can use the topological transitivity again and find an $n > m$ and a $z \in U$ with $f^n(z) \in U$. But then we have $0 < m < n$ and $z, f^n(z) \in U$ while $f^m(z) \notin U$ and this violates our earlier lemma. \square

We know now that for maps *on an interval* the only condition that has to be checked for Devaney's definition of chaos is the first one, topological transitivity. Note that the proof cannot be generalized for higher dimensions or the unit circle S^1 because our lemma uses the ordering on \mathbb{R} in an essential way.

For completeness we note that there are no other trivialities in Devaney's definition when restricted to intervals:

A continuous function on an interval whose periodic points are dense doesn't need to have sensitive dependence on initial conditions.

The identity function on any interval trivially proves this.

A continuous function on an interval which has sensitive dependence on initial conditions and whose periodic points are dense does not have to be transitive.

Define on $I = \mathbb{R}_+$ the function

$$f(x) = \begin{cases} 3x & 0 \leq x < \frac{1}{3} \\ -3x + 2 & \frac{1}{3} \leq x < \frac{2}{3} \\ 3x - 2 & \frac{2}{3} \leq x < 1 \\ f(x - 1) + 1 & x \geq 1 \end{cases}$$

It is sensitive on initial conditions since $|df/dx(x)| = 3$ for all points on I , so every neighbourhood around a point will expand under iteration. It is easy to establish that f^n has $3^n - 2$ fixed points between any two integer values with distances between these points smaller than $(\frac{1}{3})^{n-1}$, so the periodic points are dense. But since $f([0, 1]) = [0, 1]$ the function is not topologically transitive. When one restricts this function to the interval $I = [0, 2]$ one sees that it is a counterexample for finite I as well.

A continuous function on an interval which has sensitive dependence on initial conditions doesn't need to have periodic points which are dense.

As counterexample take the interval $I = [0, \frac{3}{4}]$ and the function

$$f(x) = \begin{cases} \frac{3}{2}x & 0 \leq x < \frac{1}{2} \\ \frac{3}{2}(1-x) & \frac{1}{2} \leq x \leq \frac{3}{4} \end{cases}$$

Sensitive dependence is clear again since the function is expanding, but there can be no periodic points in $]0, 3/8[$ since it is easy to establish that any trajectory with initial value in this subinterval, will not return there. For a counterexample in the infinite case, take $I = \mathbb{R}_+$ and $f(x) = 2x$.

REFERENCES

1. R. Devaney, *An Introduction to Chaotic Dynamical Systems*, Addison-Wesley, 1989.
2. J. Banks, J. Brooks, G. Cairns, G. Davis and P. Stacey, On Devaney's definition of Chaos, *American Mathematical Monthly*, 99 (1992) 332–334.
3. D. Assaf, IV and S. Gadbois, Definition of chaos, letter in *American Mathematical Monthly*, 99 (1992) 865.
4. L. S. Block and W. A. Coppel, *Dynamics in One Dimension*, Lecture Notes in Mathematics no. 1513, Springer-Verlag 1992.

*Dept. of Applied Mathematics
University of Twente
Enschede, The Netherlands*

*Department of Mathematics
Åbo Akademi
Åbo, Finland*

Proof of a Mixed Arithmetic-Mean, Geometric-Mean Inequality

Kiran Kedlaya*

The following conjecture was made by F. Holland in [3].

Conjecture. *Let x_1, x_2, \dots, x_n be positive real numbers. The arithmetic mean of the numbers*

$$x_1, \sqrt{x_1 x_2}, \sqrt[3]{x_1 x_2 x_3}, \dots, \sqrt[n]{x_1 x_2 \cdots x_n}$$

does not exceed the geometric mean of the numbers

$$x_1, \frac{x_1 + x_2}{2}, \frac{x_1 + x_2 + x_3}{3}, \dots, \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

There is equality if and only if $x_1 = x_2 = \cdots = x_n$.

*The author wishes to thank Cecil Rousseau for his help in preparing this note.

To prove this conjecture, we begin with the following bit of combinatorics. The utility of this construction will become clear as the five properties listed below are invoked. (The vectors can be explicitly constructed by trial and error up to about $n = 5$ so as to satisfy the properties; from these the general formula was deduced.)

Lemma. The vectors $\mathbf{a}(i, j) = (a_1(i, j), a_2(i, j), \dots, a_n(i, j))$ given by

$$a_k(i, j) = \binom{n-i}{j-k} \binom{i-1}{k-1} / \binom{n-1}{j-1} \quad (1)$$

$$= \frac{(n-i)!(n-j)!(i-1)!(j-1)!}{(n-1)!(k-1)!(n-i-j+k)!(i-k)!(j-k)!} \quad (2)$$

($i, j = 1, 2, \dots, n$) satisfy

- (i) $a_k(i, j) \geq 0$ for all i, j, k ,
- (ii) $a_k(i, j) = 0$ for $k > \min(i, j)$,
- (iii) $a_k(i, j) = a_k(j, i)$ for all i, j, k ,
- (iv) $\sum_{k=1}^n a_k(i, j) = 1$ for all i, j ,
- (v) $\sum_{i=1}^n a_k(i, j) = \begin{cases} n/j & \text{for } k \leq j \\ 0 & \text{for } k > j. \end{cases}$

Proof of the Lemma: Properties (i) and (ii) are obvious, and property (iii) is transparent in (2). Property (iv) is a consequence of the standard Vandermonde identity and property (v) follows from a variation of the Vandermonde identity obtained through appropriate use of upper negation. (See [1], p. 169.) Specifically,

$$\sum_{k=1}^n a_k(i, j) = \sum_{k=1}^n \binom{n-i}{j-k} \binom{i-1}{k-1} / \binom{n-1}{j-1} = 1.$$

and for $k \leq j$,

$$\sum_{i=1}^n a_k(i, j) = \sum_{i=1}^n \binom{n-i}{j-k} \binom{i-1}{k-1} / \binom{n-1}{j-1} = \binom{n}{j} / \binom{n-1}{j-1} = \frac{n}{j}.$$

Alternatively, $\binom{n-i}{j-k} \binom{i-1}{k-1}$ can be interpreted as the number of j -element subsets of $\{1, \dots, n\}$ whose k -th element is i . When summed over i this counts all j -element subsets of $\{1, \dots, n\}$ for $k \leq j$, which number $\binom{n}{j}$; when summed over k this counts only those subsets containing i , which number $\binom{n-1}{j-1}$.

Proof of the Conjecture: We use the notation of Hardy, Littlewood and Pólya [2] for weighted arithmetic and geometric means:

$$\mathcal{M}(\mathbf{x}, \mathbf{a}) = \sum_{k=1}^n a_k x_k \quad \text{and} \quad \mathcal{G}(\mathbf{x}, \mathbf{a}) = \prod_{k=1}^n x_k^{a_k},$$

where $\mathbf{a} = (a_1, a_2, \dots, a_n)$ is an n -tuple of nonnegative numbers satisfying $\sum_{k=1}^n a_k = 1$. By the arithmetic-mean, geometric-mean inequality

$$\mathcal{M}(\mathbf{x}, \mathbf{a}) \geq \mathcal{G}(\mathbf{x}, \mathbf{a}), \quad (3)$$

with equality if and only if x_k is constant over all k for which $a_k > 0$. Let $\mathcal{M}(i, j)$ and $\mathcal{G}(i, j)$ be the means obtained by setting $\mathbf{a} = \mathbf{a}(i, j)$ in $\mathcal{M}(\mathbf{x}, \mathbf{a})$ and $\mathcal{G}(\mathbf{x}, \mathbf{a})$ respectively. These are means because of properties (i) and (iv). Using properties

(ii) and (v) and the AM-GM inequality (3), we obtain

$$\frac{x_1 + x_2 + \cdots + x_j}{j} = \frac{1}{n} \sum_{k=1}^n x_k \sum_{i=1}^n a_k(i, j) = \frac{1}{n} \sum_{i=1}^n \mathcal{M}(i, j) \geq \frac{1}{n} \sum_{i=1}^n \mathcal{S}(i, j). \quad (4)$$

Taking the geometric mean of both sides over j , we have

$$\left(\prod_{j=1}^n \frac{x_1 + x_2 + \cdots + x_j}{j} \right)^{1/n} \geq \frac{1}{n} \prod_{j=1}^n \left(\sum_{i=1}^n \mathcal{S}(i, j) \right)^{1/n}. \quad (5)$$

From Hölder's inequality ([2], p. 21), we know

$$\frac{1}{n} \prod_{j=1}^n \left(\sum_{i=1}^n \mathcal{S}(i, j) \right)^{1/n} \geq \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^n \mathcal{S}(i, j)^{1/n}. \quad (6)$$

Equality holds only if every two of $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n$ are proportional where

$$\mathbf{g}_i = (\mathcal{S}(i, 1), \mathcal{S}(i, 2), \dots, \mathcal{S}(i, n)) \quad (i = 1, 2, \dots, n).$$

Since $\mathcal{S}(i, 1) = x_1$ for all i , this would imply that $\mathbf{g}_1 = \mathbf{g}_2 = \cdots = \mathbf{g}_n$. But $\mathcal{S}(i, n) = x_i$, so that would imply $x_1 = x_2 = \cdots = x_n$.

Also notice that by (iii) and (v),

$$\prod_{j=1}^n \mathcal{S}(i, j)^{1/n} = \prod_{k=1}^n \prod_{j=1}^n x_k^{a_k(i, j)/n} = \prod_{k=1}^n x_k^{1/i} = \sqrt[i]{x_1 x_2 \cdots x_i}. \quad (7)$$

Now combining (5), (6) and (7) gives

$$\left(\prod_{j=1}^n \frac{x_1 + x_2 + \cdots + x_j}{j} \right)^{1/n} \geq \frac{1}{n} \sum_{i=1}^n \sqrt[i]{x_1 x_2 \cdots x_i}, \quad (8)$$

the desired inequality. As we have shown that equality cannot occur unless $x_1 = x_2 = \cdots = x_n$ (in which case equality is obvious), the conjecture of Holland is correct. ■

REFERENCES

1. R. L. Graham, D. E. Knuth and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, Reading, 1989.
2. G. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, second edition, Cambridge University Press, Cambridge, 1951.
3. F. Holland, On a mixed arithmetic-mean, geometric-mean inequality, *Mathematics Competitions* **5** (1992), 60–64.

371 Quincy Mail Center
Harvard University
Cambridge, MA 02138-6609
kedlaya@husc.harvard.edu

UNSOLVED PROBLEMS

Edited by: **Richard Guy & Richard Nowakowski**

In this department the MONTHLY presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial or related results. Typescripts should be sent to Richard Guy, Department of Mathematics & Statistics, The University of Calgary, Alberta, Canada T2N 1N4.

ApSimon's Mints Problem

Richard Guy and Richard Nowakowski

H. ApSimon, on pages 65–76 of *Mathematical Byways in Ayling, Beeling and Ceiling*, Oxford University Press, 1984, introduces the problem of trying to identify which, if any, of a number of mints are producing a certain coin with a variant material. It is only possible to detect the difference by weighing. The correct weight of a coin is known and the small difference of the variant, if it is used, is always the same. Only two weighings are allowed, and the object is to find which mints are using the variant material, but using as few sample coins as possible.

ApSimon gives the following analysis. Suppose that there are n mints. Let C be the true weight of a coin and $C(1 + \varepsilon)$ be the weight of a coin made with the variant material. For $i = 1, \dots, n$, let a_i (b_i) be the number of coins from mint i used in the first (second) weighing and let d_i , $i = 1, \dots, n$ be zero or one according as the i th mint is using the correct or the variant material. Note that the case that all $d_i = 0$ means all the mints are using the correct material. Also, we cannot have $a_i = b_i = 0$ for any i since this would give no information about mint i .

Let the results of the two weighings be W and X , so that

$$W = \sum_{i=1}^n C(1 + d_i \varepsilon) a_i, \quad \text{and} \quad X = \sum_{i=1}^n C(1 + d_i \varepsilon) b_i$$

and

$$\left(W - \sum_{i=1}^n C a_i \right) / C = \sum_{i=1}^n d_i \varepsilon a_i \quad \text{and} \quad \left(X - \sum_{i=1}^n C b_i \right) / C = \sum_{i=1}^n d_i \varepsilon b_i$$

From the previous equations we have

$$\frac{W - \sum_{i=1}^n C a_i}{X - \sum_{i=1}^n C b_i} = \frac{\sum_{i=1}^n d_i \varepsilon a_i}{\sum_{i=1}^n d_i \varepsilon b_i} = \frac{\sum_{i=1}^n d_i a_i}{\sum_{i=1}^n d_i b_i}.$$

The lefthand side is known, so to determine which mints are providing counterfeit coins we just need to choose the a_i and b_i so that the left hand side takes on $2^n - 1$ different values for the $2^n - 1$ different possible choices of the d_i .

ApSimon showed that this was always possible. He gave the solution $b_i = 1$ and $a_i = i!$ for $i = 1, \dots, n$.

Since we require $2^n - 1$ different values, then an efficient way is to have $2^{n/2}$ different sums on top and on bottom. This means that we require $n/2$ different values for the a_i and b_i at best $0, 1, \dots, \lfloor n/2 \rfloor$ and that the total number of coins is at least $\frac{1}{2} \lfloor n/2 \rfloor \lfloor (n+2)/2 \rfloor$.

For one mint, only one coin and one weighing is necessary. For two mints it is clearly necessary and sufficient to take a coin from each mint; they may be weighed separately. For 3, 4 and 5 mints, ApSimon found that the least number of coins is 4, 8, and 15. Possible values for the a_i and b_i are $\{0, 1, 2 \text{ \& } 1, 1, 0\}$; $\{0, 1, 2, 3 \text{ \& } 1, 0, 2, 2\}$ or $\{0, 1, 1, 4 \text{ \& } 2, 0, 1, 1\}$; and $\{1, 0, 1, 4, 5 \text{ \& } 1, 2, 2, 5, 0\}$.

He asks two questions:

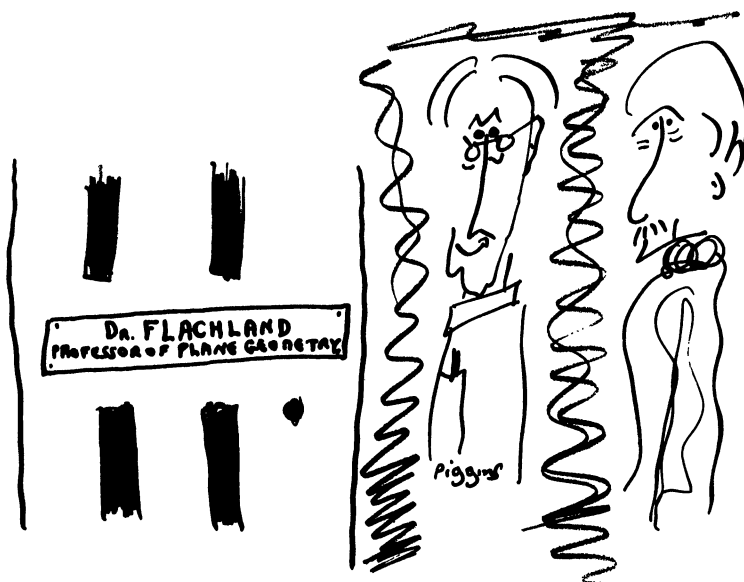
1. Are 2^{n-1} coins always sufficient?
2. What is the least number of coins needed for n mints?

It was found by computer that 38 coins suffice for 6 mints by taking the a_i to be $\{1, 1, 2, 3, 5, 8\}$ and the b_i to be $\{11, 6, 5, 3, 1, 1\}$ and that 74 coins suffice for 7 mints by taking $\{a_i\} = \{0, 1, 3, 6, 10, 15, 21\}$ and $\{b_i\} = \{1, 5, 8, 14, 10, 8, 0\}$.

A result that is better for $9 \leq n \leq 12$ than ApSimon's $\sum_{i=1}^n i!$ is obtained by noting that $(4^{n+1} - 1)/3$ coins suffice, where the a_i are $\{0, 1, 2, \dots, n-1\}$ and the b_i are $\{4, 4^2, \dots, 4^n\}$. Curiously, this fails for $n = 13$.

We can ask a further question:

3. If we are using the least number of coins, is it true that there is a choice of $\{a_i\}$ and $\{b_i\}$ such that for some j and k , $a_j > b_j$ and $a_k < b_k$?



OH! HE DOESN'T KNOW ONE END OF
A MOBIUS STRIP FROM THE OTHER.

THE AUTHORS

ED BENDER was a student of Olga Taussky-Todd at Caltech and wrote his doctoral thesis in the borderland between matrices and quadratic forms. He worked for several years as a research mathematician at the Centers for Communications Research and has been a Professor at UCSD since 1974. His primary research interest is (asymptotic) combinatorics; however, he also has interests in population biology and computer science.

OREN PATASHNIK received his bachelor's degree from Yale and his Ph.D. in computer science from Stanford, with a stint at Bell Labs in between. He has been at the Center for Communications Research since 1990. He's done research in combinatorics, graph theory, the analysis of algorithms, and positional (tic-tac-toe-like) games. He's also done time, during the past decade, on BibTEX, the bibliography processor for T_EX and L^AT_EX.

HOWARD RUMSEY, JR. received his B.S., M.S., and Ph.D. degrees in mathematics from Caltech. Over the last thirty years he has worked at MIT, JPL, Caltech, and the Centers for Communications Research, both in Princeton and La Jolla. His research interests have included combinatorics, coding theory, mapping the surface of Venus, and, more recently, symbolic computing.

RALPH WALDE received his Ph.D. from the University of California, Berkeley in 1967 where he wrote his dissertation on non-associative algebras under the direction of Robert Brown. He taught at the University of Minnesota from 1967 to 1972 and has taught at Trinity College in Hartford since then. His present research interests are dynamical systems and the theory of computation.

PAULA RUSSO received her Ph.D. from Indiana University in 1984. She wrote her dissertation on Several Complex Variables under the direction of Herbert Alexander. She taught at DePaul University, Michigan State University, and the University of Washington, before coming to Trinity College in 1987. In addition to several complex variables, she is interested in dynamical systems, particularly complex analytic dynamics.

CESAR DELGADO is Assistant Professor of Mathematics at *Universidad del Valle* in Cali, Colombia. At the same university he received his MS. (1984) in mathematics. Research in teaching mathematics has been his long-term interest.

ERNESTO ACOSTA was born in Bogotá, Colombia. He studied mathematics at *Universidad Nacional* in the same city. He moved to Cali where got his MS. degree in mathematics (1983) at *Universidad del Valle*. In 1986 he traveled to the United States and studied at Cornell University where he received his Ph.D in mathematics (1991) under Professor Leonard Gross. Now he is Associate Professor of Mathematics at *Universidad del Valle* and his research interest is analysis in infinite-dimensional manifolds.

ANTHONY (TONY) THOMPSON was born at Yorkshireman, graduated from University College London, and (in 1963) received one of the first Ph.D's from the new (then) University of Newcastle upon Tyne. Since 1966 he has taught at Dalhousie University in Halifax, N.S. with 3 years as Chairman and with sabbatical excursions to Edinburgh, Nanjing, Toronto and Freiburg. His main mathematical interest is the geometry of finite dimensional normed spaces (Minkowski spaces).

FRANK MORGAN works in minimal surfaces and studies the behavior and structure of minimizers in various dimensions and settings. His first book, *Geometric Measure Theory: a Beginner's Guide*, has just been followed by a second, *Riemannian Geometry: a Beginner's Guide*. He is currently writing a third, *Calculus in One Semester*. Morgan went to MIT and Princeton, where his thesis advisor, Fred Almgren, introduced him to minimal surfaces. He then taught for ten years at MIT, where he served for three years as Undergraduate Mathematics Chairman, received the Everett Moore Baker Award for excellence in undergraduate teaching, and held the Cecil and Ida Green Career Development Chair. He spent leave years at Rice, Stanford, and the Institute for Advanced Study. He has just received one of the first MAA national awards for distinguished teaching. Morgan now serves at Williams as Mathematics Department Chair and Codirector of an NSF undergraduate research project. In November, 1993, he was elected to the Council of the AMS. He is co-organizer of the Hudson River undergraduate mathematics conference to be held in Albany on April 9, 1994.

KAREN HUNGER PARSHALL is Associate Professor of Mathematics and History at the University of Virginia. She did her undergraduate work at the University of Virginia in mathematics and French prior to earning an M.S. in mathematics there. Under the direction of I. N. Herstein and Allen G. Debus at the University of Chicago, she completed her doctorate in history. Her research centers primarily on the historical development of algebra in the nineteenth and early twentieth centuries and on the establishment of a mathematical research community in the United States.

Jacobi's Theorem

In 1869, J. J. Sylvester gave an address to the British Association. One of the participants at that meeting was H. Jacobi, a physicist from St. Petersburg. He recounted to Sylvester the following episode concerning his celebrated brother C. G. J. Jacobi. (I have translated the French.)

“Speaking to my deceased brother, one day, on the necessity of monitoring an observation by repeated experiments—even when these verified the hypothesis—he told of having discovered a remarkable law in the Theory of Numbers—a law, which he did not in the least doubt, held in all generality. However, wanting to be exceptionally cautious, or rather, wishing to do something he considered unnecessary but prudent, he substituted a digit which he chose at random, or possibly by a kind of inspiration, and this choice did not verify the formula. Every other digit which he tried confirmed the result. Later he succeeded in proving that the digit he had chosen was one of a class which forms the only exceptions to his formula.

This curious fact stayed in my memory, but since he passed away about 30 years ago, I do not recall any details.”

Question to readers of AMM. What was Jacobi's theorem to which his brother alludes?

*Submitted by Raymond Ayoub
Department of Mathematics
Pennsylvania State University
University Park, PA 16802*

PROBLEMS AND SOLUTIONS

Edited by:

Richard T. Bumby, Fred Kochman and Douglas B. West

Proposed problems should be sent to the MONTHLY PROBLEMS address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.

Solutions of published problems should arrive before September 30, 1994 at the MONTHLY PROBLEMS address given on the inside front cover. Solutions should be typed with double spacing, including the problem number and the solver's name and mailing address. Two copies suffice. A self-addressed postcard or label should be included if an acknowledgment is desired.

*An asterisk (*) after the number of a problem, or part of a problem, indicates that no solution is currently available. Partial solutions will be useful in such cases. Otherwise, the published solution is likely to be based on a solution which is complete and correct. Of course, an elegant partial solution or a method leading to a more general result is always useful and welcome. In addition, references to other appearances of MONTHLY problems or to solutions of these problems in the literature are also solicited.*

PROBLEMS

10375. *Proposed by John Brillhart and J. S. Lomont, University of Arizona, Tucson, AZ.*

Find the complete solution of the recurrence

$$U_{n+2} = 2(2n+3)^2 U_{n+1} - 4(n+1)^2(2n+1)(2n+3)U_n \quad (n \geq 0).$$

10376. *Proposed by Nobuhisa Abe, Oita, Japan.*

Determine all integer solutions of

$$x(x+1)(x+2)(x+3)(x+4)(x+5) = y^2 - 1.$$

10377. *Proposed by Kathryn R. Laberteaux, student, University of Michigan, Ann Arbor, MI.*

On the final exam in a linear algebra class, I was asked to express the statement "A is Hermitian" in the form of a matrix identity. I should have written " $A = A^*$," but out of haste and exhaustion I wrote " $AA^* = A^2$ " instead. Was my answer correct?

10378. *Proposed by Bjorn Poonen (student), University of California, Berkeley, CA.*

Given that point D is in the interior of $\triangle ABC$ and that there are real numbers a, b, c, d such that $AB = ab$, $AC = ac$, $AD = ad$, $BC = bc$, $BD = bd$, and $CD = cd$, prove that $|\angle ABD| + |\angle ACD| = \pi/3$.

10379. *Proposed by Michael Hirsch, Emory University, Atlanta, GA, and Jonathan L. King, University of Florida, Gainesville, FL.*

Consider a two-dimensional world that three-dimensional beings would see as a vertical plane. This world is endowed with *gravity* pulling objects in the downward direction. An experiment is performed in which various convex polygonal figures are placed on a “table” (i.e., a horizontal line segment). A convex polygon P is called *stable* on an edge e if, when placed with edge e on the table and rolled slightly either way, gravity causes P to roll back to rest on e .

Let $S(P)$ denote the number of stable edges of P . What is the minimum of $S(P)$ over all convex polygons P of uniform density?

10380. *Proposed by Michael Slater, University of Bristol, Bristol, England.*

Suppose that f_1, \dots, f_n are continuous real periodic functions, and that $\sum_{i=1}^n f_i$ is a constant function, while no sum of fewer than n of the f_i is a constant function. Show that the f_i have a common period.

10381. *Proposed by Marcin E. Kuczma, University of Warsaw, Warszawa, Poland.*

Determine all real valued functions f on the integer lattice \mathbb{Z}^2 such that $f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$ for every pair of orthogonal vectors \mathbf{u}, \mathbf{v} in \mathbb{Z}^2 .

NOTES

Notes: (10378) Notation such as $|\angle ABD|$ is used for the size (always non-negative) of the angle in radians. Since D is an interior point, $|\angle ABD| + |\angle ACD|$ is *a priori* between 0 and π .

SOLUTIONS

Keeping the Ones Apart

10196 [1992, 162]. *Proposed by Barry Hayes, Stanford University, Stanford CA, and David S. Pearson, Cornell University, Ithaca NY.*

Let M_n be the set of n -bit binary strings containing no pairs of consecutive ones. For example,

$$M_3 = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0), (1, 0, 1)\}.$$

Find the probability p_n that if $(\delta_1, \delta_2, \dots, \delta_n)$ and $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ are in M_n , then

$$(\max\{\delta_1, \varepsilon_1\}, \max\{\delta_2, \varepsilon_2\}, \dots, \max\{\delta_n, \varepsilon_n\})$$

is in M_n .

Solution by Michael Andreoli, Miami Dade Community College (North), Miami, FL. The desired probability is given by

$$P_n = \frac{5 \cdot 2^{n+2}}{\sqrt{13}} \frac{(1 + \sqrt{13})^{n+2} - (1 - \sqrt{13})^{n+2}}{\left((1 + \sqrt{5})^{n+2} - (1 - \sqrt{5})^{n+2}\right)^2}.$$

An n -bit binary string may be identified with a sequence of n Bernoulli trials by letting 1 denote success and 0 failure. Consider a sequence of independent Bernoulli trials with success probability p and failure probability $q = 1 - p$. If T_n denotes the probability that there is no run of two consecutive successes in the first n trials, then conditioning on the outcome of the last trial yields.

$$T_n = qT_{n-1} + pqT_{n-2}, \quad \text{with } T_0 = T_1 = 1. \quad (1)$$

Now suppose two players simultaneously and independently toss fair coins. Say that a win occurs on the k th toss if at least one player has a head on that toss. Consider the events.

$A_n = \{\text{Player I has no run of two consecutive heads in the first } n \text{ tosses}\}$

$B_n = \{\text{Player II has no run of two consecutive heads in the first } n \text{ tosses}\}$

$C_n = \{\text{There is no run of two consecutive wins in the first } n \text{ tosses}\}.$

We then have

$$p_n = P(C_n | A_n \cap B_n) = \frac{P(A_n \cap B_n \cap C_n)}{P(A_n \cap B_n)} = \frac{P(C_n)}{P(A_n \cap B_n)}. \quad (2)$$

Letting $Q_n = P(A_n) = P(B_n)$ and applying (1) with $p = 1/2$, we obtain

$$Q_n = \frac{1}{2}Q_{n-1} + \frac{1}{4}Q_{n-2}, \quad \text{with } Q_0 = Q_1 = 1,$$

which has the solution

$$Q_n = \left(\frac{\sqrt{5} + 3}{2\sqrt{5}} \right) \left(\frac{1 + \sqrt{5}}{4} \right)^n + \left(\frac{\sqrt{5} - 3}{2\sqrt{5}} \right) \left(\frac{1 - \sqrt{5}}{4} \right)^n.$$

Similarly, letting $R_n = P(C_n)$ and applying (1) with $p = 1/4$, we obtain

$$R_n = \frac{1}{4}R_{n-1} + \frac{3}{16}R_{n-2}, \quad \text{with } R_0 = R_1 = 1,$$

which has the solution

$$R_n = \left(\frac{\sqrt{13} + 7}{2\sqrt{13}} \right) \left(\frac{1 + \sqrt{13}}{8} \right)^n + \left(\frac{\sqrt{13} - 7}{2\sqrt{13}} \right) \left(\frac{1 - \sqrt{13}}{8} \right)^n.$$

Substituting these values for Q_n and R_n into (2) and simplifying, we obtain the result stated at the outset.

Editorial comment. Most solvers first observed that $|M_n|$ is the $n + 2^{nd}$ Fibonacci number (see, e.g., Example 10.10 in Ralph P. Grimaldi, *Discrete and Combinatorial Mathematics*, 2nd ed., Addison-Wesley, 1989). They then used a similar counting argument to establish that the number of pairs from M_n whose bit-wise or is in M_n satisfies the recursion $G_{n+1} = G_n + 3G_{n-1}$. The result follows easily by standard techniques. A similar counting argument was exploited in Richard Austin and Richard Guy, "Binary sequences without isolated ones," *Fibonacci Quarterly* 16 (1978), 84–86.

Solved also by J. C. Binz (Switzerland), W. J. Buhler (Germany), D. Callan, R. J. Chapman (U.K.), W. Y. C. Chen, C. P. Grant, I. Kastanas, K. S. Kedlaya (student), K. M. Levasseur, J. H. Lindsey II, G.A. Martin, T. S. Norfolk, H. Schmidt Jr., M. Stamp, J. H. Steelman, R. Stong, F. B. Strauss, C. Subi, D. White, National Security Agency Problems Group, Theory First, the 1993 UK IMO team (students, U.K.), and the proposers.

The Longest Shortest Closed Walk

10204 [1992, 265–6]. *Proposed by Edgar A. Ramos and Douglas B. West, University of Illinois, Urbana, IL.*

Given a strongly connected directed graph G , let $s(G)$ be the length of the shortest closed walk visiting every vertex. Determine, for each positive integer n , the maximum value of $s(G)$ over strongly connected directed graphs with n vertices.

Solution by David M. Wells, Penn State University, New Kensington, PA. The maximum value is $\lfloor (n + 1)^2/4 \rfloor$. To prove the upper bound, let G be a strongly connected digraph with vertices v_1, \dots, v_n . Let $d(i, j)$ denote the length of the shortest directed path from v_i to v_j , and let k denote the maximum $d(i, j)$ over all $i \neq j$. We may index the vertices so that this longest shortest path is v_1, \dots, v_{k+1} . This walk can be extended to a closed walk by appending paths to visit vertices v_{k+2}, \dots, v_n and v_1 in order. Each of these segments has length at most k , so we have constructed a spanning closed walk of length at most $k(n + 1 - k)$. For $1 \leq k \leq n$, this quantity is bounded by $\lfloor (n + 1)^2/4 \rfloor$.

Letting $k = \lfloor (n + 1)/2 \rfloor$, the maximum is attained by the digraph consisting of edges $v_i v_{i+1}$ for $1 \leq i \leq k - 1$, edges $v_k v_j$ for $k + 1 \leq j \leq n$, and edges $v_j v_1$ for $k + 1 \leq j \leq n$.

Editorial comment. Most solvers gave a similar proof, starting with a longest path or a longest cycle and building to a spanning walk by including the remaining vertices. Some solvers began with a shortest spanning closed walk and partitioned it into cycles that each contain a vertex not on any other cycle. Hence the total length is at most $k(n + 1 - k)$, where k is the number of cycles in the decomposition.

Paul Cull noted that he published this result in his article "Tours of graphs, digraphs, and sequential machines," *IEEE Trans. Computers* C29(1980), 50–54, using the proof printed here. The solution of S. L. Hakimi and E. F. Schmeichel pointed out that the related problem in which the walk is not required to be closed has the solution $\lfloor n^2/4 \rfloor$.

Solved also by P. Cull, J. W. Grossman, S. L. Hakimi & E. F. Schmeichel, R. High, G. Johns & T. Sprague & T. Sipka, K. S. Kedlaya (student), N. Komanda, O. P. Lossers (The Netherlands), R. F. McCoart, Jr., E. R. Scheinerman, R. Stong, and the proposer.

Spectral Representation of the Laplace Transform

10237 [1992, 571]. *Proposed by Paul R. Chernoff, University of California, Berkeley, CA.*

Consider the Laplace transform \mathcal{L} as an operator on $L^2(0, \infty)$. Show that \mathcal{L} is a bounded self-adjoint operator which is unitarily equivalent to the “position operator” X = multiplication by the coordinate x on $L^2(-\sqrt{\pi}, \sqrt{\pi})$.

Solution by the proposer. It is convenient to work with the measure dt/t , which is the Haar measure for the multiplicative group \mathbb{R}_+ of positive reals. Accordingly, define the unitary operator $W: L^2(0, \infty) \rightarrow L^2(\mathbb{R}_+; dt/t)$ by $(Wf)(t) = f(t)\sqrt{t}$. Then define $L_1 = WLW^{-1}$ on $L^2(\mathbb{R}_+; dt/t)$. We have the formula

$$\begin{aligned}(L_1\phi)(x) &= \sqrt{x} \int_0^\infty e^{-xt} \frac{\phi(t)}{\sqrt{t}} dt \\ &= \int_0^\infty \sqrt{xt} e^{-xt} \phi(t) \frac{dt}{t} \\ &= \int_0^\infty \sqrt{t} e^{-t} \phi\left(\frac{t}{x}\right) \frac{dt}{t} \\ &= \int_0^\infty \sqrt{t} e^{-t} \tilde{\phi}\left(\frac{x}{t}\right) \frac{dt}{t}\end{aligned}$$

Here $\tilde{\phi}(t) = \phi(t^{-1})$. Note that the map $\phi \mapsto \tilde{\phi}$ is unitary on $L^2(\mathbb{R}_+; dt/t)$.

Thus $L_1\phi = k * \tilde{\phi}$, where $k(t) = \sqrt{t} e^{-t}$ and $*$ is convolution on the multiplicative group \mathbb{R}_+ . Since k is an L^1 function with respect to the Haar measure dt/t , it follows that L_1 is a bounded operator on $L^2(\mathbb{R}_+; dt/t)$, and therefore the unitarily equivalent operator L is bounded on $L^2(0, \infty)$. Therefore L is a self-adjoint Carleman integral operator.

Next, via the Mellin transform (the Fourier transform on the group \mathbb{R}_+), we see that L_1 is unitarily equivalent to the operator N on $L^2(-\infty, \infty)$ given by the formula

$$(N\psi)(s) = K(s)\psi(-s)$$

where K is the Mellin transform of k :

$$K(s) = \int_0^\infty t^{2\pi is} \sqrt{t} e^{-t} \frac{dt}{t} = \Gamma(1/2 + 2\pi is).$$

Thus $N = M_K R$ where M_K is multiplication by K and R is the unitary reflection operator $(R\psi)(s) = \psi(-s)$. Note the relation $K(-s) = \Gamma(1/2 - 2\pi is) = \overline{K(s)}$, which is precisely equivalent to the self-adjointness of the operator N .

Define $\omega(s) = K(s)/|K(s)|$. Then $|\omega(s)| = 1$ and $\omega(-s) = \overline{\omega(s)} = \omega(s)^{-1}$. Let

$$\begin{aligned}H^+ &= \{\psi \in L^2(-\infty, \infty) : \psi(-s) = \omega(s)^{-1}\psi(s)\} \\ H^- &= \{\psi \in L^2(-\infty, \infty) : \psi(-s) = -\omega(s)^{-1}\psi(s)\}.\end{aligned}$$

Then it is easy to see that L^2 is the orthogonal direct sum of the closed subspaces H^+ and H^- . (Indeed, an arbitrary $\psi \in L^2$ decomposes as $\psi^+ + \psi^-$ where $\psi^\pm(s) = (\psi(s) \pm \omega(s)\psi(-s))/2$.)

The subspaces H^\pm are invariant under the operator N . For example, if $\psi \in H^+$ then

$$(N\psi)(s) = K(s)\psi(-s) = K(s)\omega(s)^{-1}\psi(s) = |K(s)|\psi(s)$$

and $|K|\psi \in H^+$ because $|K|$ is an even function. Similarly, if $\psi \in H^-$, we have $(N\psi)(s) = -|K(s)|\psi(s)$.

Finally, note that the map $\psi \mapsto \sqrt{2} \cdot \psi$ restricted to $(0, \infty)$ is unitary from H^\pm to $L^2(0, \infty)$. Conclusion: N^+ (the restriction of N to H^+) is unitarily equivalent to the multiplication operator $M_{|k|}$ on $L^2(0, \infty)$. Likewise N^- is unitarily equivalent to $M_{-|k|}$. Of course N is the direct sum of N^+ and N^- .

Now

$$|K(s)| = \left(\Gamma\left(\frac{1}{2} + 2\pi is\right) \Gamma\left(\frac{1}{2} - 2\pi is\right) \right)^{1/2} = \left(\frac{\pi}{\cosh(2\pi^2 s)} \right)^{1/2},$$

a monotone function on $(0, \infty)$ with $0 < |K| < \sqrt{\pi}$. Accordingly $M_{|K|}$ is unitarily equivalent to the position operator X on $L^2(0, \sqrt{\pi})$, and $M_{-|K|}$ is unitarily equivalent to the position operator on $L^2(-\sqrt{\pi}, 0)$.

To sum up: by putting together our chain of unitary equivalences, we have proved that the Laplace transform operator L is self-adjoint, bounded (so a Carleman integral operator), and unitarily equivalent to the position operator X on $L^2(-\sqrt{\pi}, \sqrt{\pi})$. (In particular $\|L\| = \sqrt{\pi}$, a known fact; cf. the discussion of Hardy-type inequalities in Nelson Dunford and Jacob T. Schwartz, *Linear Operators I*, Interscience, 1958.)

Editorial comment. Jerome Goldstein supplied a reference to D. S. Gilliam, J. R. Schulenberger and J. L. Lund, "Spectral representation of the Laplace and Stieltjes transforms," *Mat. Applic.* 7 (1988), 101–107. This paper contains a solution of this problem, which the authors characterize as "at least partially known," and they also point out that this provides a rare example of a bounded operator with a continuous spectrum for which the spectral representation is explicitly known.

Solved also by R. J. Chapman (U.K.) and A. Vogt.

Moments of a Bivariate Normal Distribution

10246 [1992, 676]. *Proposed by B. C. Carlson, Iowa State University, Ames, IA.*

For integers m and n with $n \geq 0$ and $-n \leq m \leq n$, find values of A , N , and λ such that

$$\frac{1}{2\pi(1-\rho^2)^{1/2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^{n+m} y^{n-m} \exp\left(\frac{-x^2 + 2\rho xy - y^2}{2(1-\rho^2)}\right) dx dy = AC_N^\lambda(-\rho)$$

for $-1 < \rho < 1$, where C_N^λ is a Gegenbauer polynomial.

Solution by F. W. Steutel, Eindhoven University of Technology, Eindhoven, The Netherlands. Clearly, the integral to be evaluated, denoted here by $I_{m,n}$, is given by

$$I_{m,n} = \mathbf{E} X^{n+m} Y^{n-m},$$

where (X, Y) is a normally distributed random vector with zero means, unit variances and correlation coefficient ρ , and where \mathbf{E} denotes expectation. As we have

$$\sum_{m=-n}^n \binom{2n}{n+m} X^{n+m} Y^{n-m} t^{n-m} = (X + tY)^{2n},$$

it follows that

$$I_{n,m} = \left(\binom{2n}{n+m} \right)^{-1} \text{coefficient of } t^{n-m} \text{ in } E(X + tY)^{2n}.$$

Since $X + tY$ is a normally distributed random variable with mean zero and variance $1 + 2\rho t + t^2$ we have

$$E(X + tY)^{2n} = 1 \cdot 3 \cdot 5 \cdots (2n - 1)(1 + 2\rho t + t^2)^n.$$

From the definition of $C_N^\lambda(z)$ we then obtain

$$I_{n,m} = I_{n,-m} = \frac{(n+m)!(n-m)!}{2^n n!} C_{n-|m|}^{-n}(-\rho).$$

Solved also by J. Anglesio (France), R. J. Chapman (U.K.), I. Kastanas, O. P. Lossers (The Netherlands), H. J. Seiffert (Germany), and the proposer.

Collaborating editors: *David F. Appleyard, Paul T. Bateman, Duane M. Broline, Barry W. Brunson, Frank S. Cater, Gulbank D. Charkarian, Underwood Dudley, Gerald A. Edgar, Michael A. Filaseta, Ira M. Gessel, Richard A. Gibbs, Jerrold R. Griggs, Douglas A. Hensley, John R. Isbell, Mourad E. H. Ismail, Murray Klamkin, Daniel J. Kleitman, Frederick W. Luttman, Frank B. Miles, Richard Pfiefer, Stephen L. Portnoy, J. O. Shallit, John Henry Steelman, Kenneth B. Stolarsky, David E. Tepper, Douglas B. Tyler, Daniel Ullman, and William E. Watkins.*

I know, indeed, and can conceive of no pursuit so antagonistic to the cultivation of the oratorical faculty ... as the study of Mathematics. An eloquent mathematician must, from the nature of things, ever remain as rare a phenomenon as a talking fish, and it is certain that the more anyone gives himself up to the study of oratorical effect the less will he find himself in a fit state to mathematicize. It is the constant aim of the mathematician to reduce all his expressions to their lowest terms, to retrench every superfluous word and phrase, and to condense the Maximum of meaning into the Minimum of language. He has to turn his eye ever inwards, to see everything in its dryest light, to train and inure himself to a habit of internal and impersonal reflection and elaboration of abstract thought, which makes it most difficult for him to touch or enlarge upon any of those themes which appeal to the emotional nature of his fellow-men. When called upon to speak in public he feels as a man might do who has passed all his life in peering through a microscope, and is suddenly called upon to take charge of an astronomical observatory. He has to get out of himself, as it were, and change the habitual focus of his vision.

—J.J. Sylvester

REVIEWS

Edited by **Darrell Haile**

Indiana University, Bloomington, IN 47405

Mathematics of the 19th Century: Mathematical Logic, Algebra, Number Theory, Probability Theory. A. N. Kolmogorov and A. P. Yushkevich, ed. Basel/Boston/Berlin: Birkhäuser Verlag, 1992. xiv + 308 pp. Hardcover. \$149.50.

Reviewed by **Karen Hunger Parshall**

The history of mathematics poses many difficulties for those who choose to pursue it. Who comprises the audience for the subject? Mathematicians? Historians of mathematics? Historians of science? Historians? Some of the above? All of the above? With what questions and issues should the subject deal? Who did what when, and who did what first? What theorems led up to theorem *Y*, and how were they proven? In what ways did the social and political milieu of the country in which mathematician *X* lived affect the way in which theorem *Y* was conceived or proven? How did the particular biographical details of mathematician *X* influence mathematics as he or she chose to pursue it? How did the broader philosophical, intellectual, and cultural trends of the time influence the mathematical theories produced? Should the history of mathematics address all of these questions? Some of these questions? Should any single study in the history of mathematics try to answer all of these—and other possible—questions simultaneously? Or should distinct studies treat these various concerns separately?

The issue of audience is a sticky one. Theoretically speaking, the answer to the first multiple choice question above should be “all of the above.” The history of mathematics should make mathematicians more conscious of their roots and inform them of the rich and complex past of their chosen field. For obvious reasons, it should interest historians of mathematics themselves. Historians of science, regardless of the particular science or time frame they study, should benefit from the added dimension that an understanding of concurrent mathematical developments would bring to their work. And, similarly, historians, in order to understand fully the period defining their immediate research interests, should be aware of its scientific and mathematical matrix.

This is, of course, just speaking theoretically. In practice, each possible constituency has its own notion of how the subject should be done. “Historians should write history and not mathematics,” the mathematician might say, and the historians—and even some historians of science—might agree. Or the mathematician might just as likely charge the historian to “tell us where the theorems came from” and to “forget all of the historical mumbo-jumbo about politics and philosophy.” Here, however, those who find technical mathematical discussions less than illuminating would hardly concur. One mathematician, John Ewing, has even argued that the historian of mathematics should write “poetic history—the kind that Homer wrote—and the kind that many mathematicians want to read” [2,

p. 93]. In his view, the history of mathematics should satisfy the emotional needs of the mathematician seeking to appreciate and understand “both the heros and the villians” of the past [2, p. 93]. It should focus more on telling a good story, in emulation of the “history” of Eric Temple Bell [1], and less on analyzing the “dull details of a 100-year-old paper” [2, p. 94]. This implied characterization of good history would certainly not attract many—if any—adherents within the historical community. Since the history of mathematics *is*, after all, history, it cannot ignore the prescripts of its parent discipline any more than it can avoid the fact that it deals with mathematics.

The problematic nature of audience immediately suggests the difficulties surrounding the second set of questions posed above, those involving content and focus. Because the way in which historians ply their trade changes over time, what are deemed interesting and appropriate questions for historical investigation also experience historical change. This implies, moreover, that the historical standards of just a few decades ago may no longer obtain. Compare, for example, the historical climate of 1972, when Morris Kline published his mammoth book, *Mathematical Thought from Ancient to Modern Times*, with that of today [3]. So-called internalist practices in the history of science still held sway within the subdiscipline of the history of mathematics. The internal aspects of mathematics—the theorems along with their development and extension in the published works of mathematicians—defined the accepted scope of historical inquiry. Factors beyond the bounds of the technical evolution of the subject received scant—or, more likely, no—attention. Today, however, historians of mathematics try increasingly to respond to trends in scholarship both in the history of science and in history in general. While continuing to recognize the value of securing an historical grasp of internal mathematical developments, they also appreciate the important insights that an analysis of external social, intellectual, political, philosophical, and cultural factors can provide into their science. This heightened awareness and its concomitant widening of the historical perspective on mathematics have significantly deepened our understanding of the subject. Books like Herbert Mehrtens’s *Moderne, Sprache, Mathematik* [5] and Joan Richards’s *Mathematical Visions: The Pursuit of Geometry in Victorian England* [6], each in its own very different way, reflect these changing historiographical standards.

These observations now bring us to the book under review. Originally published in the former Soviet Union in 1978, *Mathematics of the 19th Century* represented the fruits of an ongoing, collaborative research program, which had already produced the three-volume *History of Mathematics from Antiquity to the Early Nineteenth Century* between 1970 and 1972 [7] and which was scheduled to generate three more volumes on various aspects of the history of nineteenth-century mathematics in addition to two volumes on the twentieth century. In their preface, this book’s editors, A. N. Kolmogorov and A. P. Yushkevich, stated their particular volume’s general objectives. It sought to “consider the development of mathematics not simply as the process of perfecting concepts and techniques for studying real-world spatial forms and quantitative relationships but as a social process as well” [p. ix]. It also aimed to “examine the interaction of mathematics with the social structure, technology, the natural sciences, and philosophy” [p. ix]. In other words, this book, a product of the late 1970’s, was intended to address both internalist and externalist issues as they pertain to the history of nineteenth-century mathematics in an effort both to understand mathematics technically and to situate it within a broader cultural framework. Had it achieved its stated goals,

it would have been both path-breaking for its day and remarkably timely in its presentation in an English translation in the early 1990's. Instead, however, this book provides a traditional, internalist account of the technical development of four areas of mathematics—mathematical logic, algebra, number theory, and probability—not during the whole of the nineteenth century, as the title implies, but rather from roughly 1800 to the more “natural chronological boundary” [p. xiii] of the 1870's or 1880's. Moreover, as the editors admit, their book “is more a collection of essays than a connected history of nineteenth . . . -century mathematics” [p. x]. As such, it provides the reader with historical glimpses—and these of varying quality—of selected areas as opposed to an integrated discussion of the history of mathematics in the nineteenth century.

The first essay treats of mathematical logic and is the least satisfying in the collection. By far the shortest chapter in the book (only thirty-three pages), it also opens not around 1800, in keeping with the book's prescribed time frame, but with the work of Aristotle in the fourth century B.C.! In two brief pages, it skims through some 2000 years to focus (for six pages) on the work of Gottfried Wilhelm Leibniz before reaching the nineteenth century and the ideas of Augustus DeMorgan, George Boole, William Stanley Jevons, John Venn, Ernst Schröder, and finally P. S. Poretskiĭ. This hyper-extended chronology almost necessarily results in an overly superficial and highly disjointed historical treatment of the subject matter at hand. Furthermore, in discussing the work of the nineteenth-century figures, the author, Z. A. Kuzicheva, merely presents brief summaries of the results in a chosen work rather than providing any meaningful historical—or mathematical—context for those results. This is not even good internalist history, much less the kind of historical approach promised in the preface, and that is really too bad because this chapter had much to offer. In particular, it touches on, but unfortunately does not explore, issues like the reception of Jevons's ideas in Russia [p. 24] and the transmission of ideas from Great Britain to the Continent and to Russia [pp. 27–31]. Had the answers to these and related questions been fleshed out, this chapter would have not only shed real light on the evolution of mathematical logic in Russia but also satisfied, at least to some extent, the editors' stated goal of viewing nineteenth-century mathematics through a wider historical lens.

The second chapter, on algebra and algebraic number theory by I. G. Bashmakova and A. N. Rudakov (with the assistance of A. N. Parshin and E. I. Slavutin), contrasts markedly with the first in both range and depth. In 100 pages, the authors survey such topics as the theory of equations, group theory, linear algebra, the theory of nonassociative and associative algebras, invariant theory, and the theories of quadratic forms and ideals, while they focus more specifically on the work of figures like Carl Friedrich Gauss, Niels Henrik Abel, Évariste Galois, William Rowan Hamilton, Ernst Kummer, Richard Dedekind, and Leopold Kronecker. Theirs is a difficult task given that algebra, as we now conceive of it, did not exist in the nineteenth century. The authors of this chapter thus attempt to tie together the various threads which ultimately interwove to form the subject. In so doing, however, they, more often than not, reach back into the nineteenth century from their contemporary point of view and “find” precursors to modern notions instead of analyzing the historical setting on its own terms. The chapter is peppered with statement like “Gauss' study of forms of the same discriminant was in effect a study of the fundamental properties of cyclic and general abelian groups” [p. 37] and “What Gauss wants to say is that the group he is considering is not cyclic but is a direct sum of two or more cyclic groups [p. 65] and “To this end

he [E.I. Zolotarev] constructed in effect the semilocal ring of p -integral numbers in $\mathbb{Q}(\theta)$ " [p. 111]. While it is mathematically the case that what Gauss and Zolotarev were doing can be expressed in these modern-day terms, to describe their work in this way with no further qualification suggests that Gauss thought in terms of groups and Zolotarev understood semilocal rings. These are value-laden terms to contemporary mathematicians and, as such, convey a skewed picture of the mathematics of the historical figures. True, the historian needs to communicate with the mathematician in understandable terms, but the changed historical climate also requires that that communication nevertheless reflect a sensitivity to the historical setting. In short, the authors of this, and indeed, of all of the chapters of this book, have quite unavoidably adopted the historiographical standards of the time in which they were initially writing, but that tends to render their work somewhat anachronistic today.

This almost inevitable aspect of the chapter and of the book aside, the discussion of nineteenth-century algebra brings quite a lot of material together in one manageable essay. In particular, it provides for the non-Russian reader an account of Zolotarev's work in the theory of integral and p -integral numbers. It also briefly and tantalizingly touches on—but does not delve into—the cultural and mathematical interaction between Zolotarev and the European mathematical community during Zolotarev's trips abroad in 1872 and 1876 [pp. 108–109]. More analysis of the effects of these trips on Zolotarev's process of mathematical maturation would have significantly deepened the reader's understanding of the history of Russian mathematics, of the relationship between that history and parallel developments in western Europe, and of the dissemination of mathematical theories across political boundaries. The chapter would also have benefitted from a well-developed conclusion which attempted some synthesis—around the notion of the establishment of an *area* called algebra—of the different sets of ideas discussed.

Like the second chapter, the third represents a substantial study of a panoply of topics which can be unified under one rubric, in this case, number theory. Written by E. P. Ozhigova (with the assistance of A. P. Yushkevich), this essay focuses particularly on the arithmetic theory of quadratic forms, the geometry of numbers, analytic methods in number theory, and the problem of transcendental numbers, and basically proceeds as a series of synopses of important works in the various topical categories. Here, the reader sees the ideas of some of the key European figures—Charles Hermite, Henry J. S. Smith, Hermann Minkowski, Peter Lejeune-Dirichlet, Bernhard Riemann, and Joseph Liouville—juxtaposed with the roughly contemporaneous researches of the Russians, A. N. Korkin, Zolotarev, A. A. Markov, G. F. Voronoï, P. L. Chebyshev, and N. V. Bugaev, among others. Although the chapter rarely fleshes out meaningful historical connections between the Russians and the Europeans, it does provide insights into the lesser known (at least among non-Russian readers) work of key Russian mathematicians. In fact, this chapter—and, indeed, this volume—is at its best in its incorporation of important Russian mathematical contributions into what have come to be recognized as mainstream historical developments in western Europe. Aside from occasional mentions of Chebyshev and Markov, most non-Russian histories which deal with nineteenth-century mathematics tend to ignore major Russian contributors because of their lack of influence outside of Russian-speaking regions. The reader of this volume, however, will come away with a heightened appreciation of the Russian mathematical endeavor in the nineteenth century.

Perhaps precisely because of this, though, that same reader will probably also come away wanting to know more than is provided about the Russian mathematical scene. In Chapter 3, for example, numerous references to St. Petersburg appear: Zolotarev and Markov earned their Master's Degrees there in 1869 and 1880, respectively [pp. 144 and 151]; Korkin held a professorship there from 1868 until his death in 1908 [p. 145]; Voronoï first studied and then served on the faculty there [p. 166]; and, in fact, the theory of algebraic numbers occupied the attention of "many mathematicians in St. Petersburg" in the 1890's [p. 166]. All of these references lead the reader to wonder what St. Petersburg was like as an academic institution and as an educational environment. How did so many associated with it come to be interested in algebraic number theory? Does it make sense to speak of a St. Petersburg "school" of algebraic number theory? If so, what characterizes its approach to the subject? If not, why not? The very short (half-page-long) conclusion to the chapter only piques the reader's curiosity further on these and related issues. "The sharply increased role of number theory," the authors state, "is attested by the long list of names of outstanding mathematicians who studied it in in the 19th century, as well as the appearance of scientific schools, of which the St. Petersburg school of number theory is a remarkable example" [p. 209]. The presentation up to this point has certainly and rightly suggested the presence of such a school in St. Petersburg, but the authors have provided neither an historical argument for it nor an analysis of it. Again, this is the sort of social historical discussion the preface had led the reader to expect, and its omission weakens the chapter's overall effectiveness.

The fourth and final chapter, by B. V. Gnedenko and O. B. Sheĭnin, does address the issue of a mathematical school at St. Petersburg, namely, the school of probability begun by B. Ya. Bunyakovskii, driven by Chebyshev, and perpetuated by Chebyshev's student, Markov. Moreover, it provides more historical indications of the mathematical influences on that school than do the preceding chapters. Here, the authors trace the evolution of probability in France from Laplace through Poisson and Cauchy, and they link that discussion to their presentation of the St. Petersburg school with the remark that "the probability-theoretic studies of the Paris school of mathematicians...soon became known and were further developed in Russia" [p. 247]. They also make passing mention of "the powerful influence of Laplace and Poisson" on Bunyakovskii [p. 248] and to Chebyshev's references to "the celebrated Geometer," Poisson, and his works [p. 256], in order to reinforce these intellectual connections.

Despite these sorts of historical linkages, however, the material presented here and elsewhere in this book tends to be treated synoptically and with a sense of tracing the prehistory of some contemporary idea or theory, rather than in historical context. Thus, the reader is told that the form of the integral limit theorem which Laplace gave in his *Théorie analytique des Probabilités* of 1812 [4] "may be considered the origin of later studies that found completion in S. N. Bernshteĭn's paper" of 1943 [p. 216] and that "Laplace's reasoning [on a problem applicable to election procedures] may be seen to belong to the prehistory of rank correlation and of the statistics of random processes" [p. 219]. This sort of "result tracing" provides a certain amount of useful information, but it generally fails to illuminate the historical figure's approach to or understanding of his or her work. It, too, reflects an older historiographical standard.

From the perspective of 1994, then, *Mathematics of the 19th Century* suffers from various shortcomings. There are also a few mistakes which, if they existed in

the original Russian text, should certainly have been corrected here. For example, the Irish mathematician, George Salmon, is said to have done his work “in England” [p. 80] and the definition of the Jacobian appears incorrectly [p. 82]. Moreover, the book’s princely price tag of \$149.50 hardly warrants the inordinate number of typographical errors and the less-than-felicitous English exposition which appears in more than several places, the latter despite the efforts of a talented team of translators which included Abe Shenitzer, H. Grant, O. B. Sheĭnin, and A. B. Aries. Taking all of this into account, however, the reader will still find much valuable technical information on four areas of the history of nineteenth-century mathematics in this volume.

REFERENCES

1. Eric Temple Bell, *Men of Mathematics*, Simon & Schuster, New York, 1937.
2. John Ewing, Essay Review of *The History of Modern Mathematics*, *Historia Mathematica* 19 (1992):93–98.
3. Morris Kline, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, New York, 1972.
4. Pierre Simon de Laplace, *Théorie analytique des Probabilités*, Courcier, Paris, 1812.
5. Herbert Mehrtens, *Moderne, Sprache, Mathematik*, Suhrkamp Verlag, Frankfurt-am-Main, 1990.
6. Joan Richards, *Mathematical Visions: The Pursuit of Geometry in Victorian England*, Academic Press, Boston, 1988.
7. A. P. Yushkevich, ed., *History of Mathematics from Antiquity to the Early Nineteenth Century* [in Russian], 3 vols., Nauka, Moscow, 1970–1972.

Departments of Mathematics and History
University of Virginia
Mathematics/Astronomy Building
Charlottesville, VA 22903-3199

Calculus Gems: Brief Lives and Memorable Mathematics. By George F. Simmons, McGraw-Hill, Inc., New York, xiv + 355 pp.

Reviewed by **David J. Pengelley**

Today we see the beginning of a flood of new publications and curricular programs aimed at ‘reforming’ how we teach calculus, rightly spurred by disenchantment with the status quo. While there is considerable diversity in this flood, there is also an emerging nationwide convergence toward common themes for reform. *Calculus Gems*, however, does not fit conveniently into the mold and buzz words of these themes. It is instead a needed, albeit likely unintended, challenge to ‘reformed’ and ‘unreformed’ alike. I hope that the congealing orthodoxy of calculus reform will be receptive to the enriching and inspiring vantage point this book provides. In fact I am part of that orthodoxy—I have been one of the principals involved in a major calculus ‘reform’ project from its inception.

What are some of the laudable features of revised calculus programs? We now train students in collaborative learning and have them actively involved in the classroom. We provide alternatives to lecturing, and emphasize writing as a way to learn. We seek a ‘leaner and livelier’ syllabus with more concentration on the

understanding of concepts and on meaningful and challenging applications, and less emphasis on standard calculational drill, especially things done better by calculator or computer. We expect students to become more active learners both in and out of class, tackling larger projects, individually or in groups, and acquiring and melding the tools of calculus to solve multi-step or open-ended problems. And we recognize the usefulness of calculators and computers as new experimental or visual tools.

Yet a fundamental spark is still missing from these aims—precisely that which often inspires us as mathematicians. We are motivated by seeing mathematics yield beautiful, often surprising results, introducing new mysteries even as it illuminates old ones; and when it fascinates and captures our imagination. Moreover, as each of us follows and perhaps contributes to the great mathematical discoveries of our day, are we not crucially stimulated by the fabric of human, philosophical and scientific interconnections involved: people and places, flows of ideas between groups and individuals, and lively interplay with other ideas in mathematics, science, and philosophy?

Do our students feel this kind of excitement when we exhort them that the fundamental theorem of calculus, the ideas it connects, and its consequences, are some of the greatest discoveries of the human species, rather than merely an eminently practical calculational device? How can we possibly convince them of this in an age when they ‘know’ that areas, volumes, graphs, etc., are available at the push of a few buttons, thanks to the brute calculating power of the digital computer? Do our students gain any sense of profundity or beauty lurking in the connections between infinite processes and numbers like π and e ? Do they even see any importance in a distinction between these numbers and their calculator approximations?

It is almost as if we hide the crown jewels of the infinite from our students, so caught up are we in imparting training for routine, technocratic applications. If we expect students to appreciate the beauty and power of calculus, we must immerse them in those very aspects which spark the imagination and delight us with a sense of awe. Perhaps because calculus is a mature subject we have forgotten that such inspiring jewels abound. They were discovered by some of the world’s greatest minds, and the context of their discovery is as rich (possibly more so) in human, scientific, and philosophical controversy and interest as that of present day science. We and our students should dare to play with them and struggle with their history and import.

Turning with this perspective to the work under review, *Calculus Gems: Brief Lives and Memorable Mathematics*, let us begin with what author George Simmons tells us about his book. “My overall aims are bound up with the question, ‘What is mathematics for?’ and with its inevitable answer, ‘To delight the mind and help us understand the world.’ I hold the logically impeccable view that there are only two kinds of students . . . : those who are attracted to mathematics; and those who are not yet attracted, but might be. My intended audience embraces both types.”

The book is in two parts, adapted from two massive appendices to his *Calculus with Analytic Geometry* textbook. The first, “entitled Brief Lives, amounts to a biographical history of mathematics from the earliest times to the late nineteenth century,” with two main purposes, first “to ‘humanize’ the subject, to make it transparently clear that great human beings created it by great efforts of genius, and thereby to increase students’ interest [A] tiny minority of men and women are drawn irresistibly to problems: their minds embrace them lovingly and wrestle

with them tirelessly until they yield their secrets. It is these who have taught the rest of us most of what we know and can do, I have written about some of these people from our past in the hope of encouraging a few in the next generation.” The second purpose is “connected with the fact that many students from the humanities and social sciences are compelled against their will to study calculus The profound connections that join mathematics to the history of philosophy, and also to the broader intellectual and social history of Western civilization, are often capable of arousing the passionate interest of these otherwise indifferent students.”

Introducing the second part of his book, entitled *Memorable Mathematics*, Simmons writes: “. . . I have collected a considerable number of miscellaneous topics [‘nuggets’] from number theory, geometry, science, etc., which I have used for the purpose of opening doors and forging links with other subjects . . . and also for breaking the routine and lifting the spirits. . . in the hope of making a few more converts to the view that mathematics, while sometimes rather dull and routine, can often be supremely interesting.” A final aspect of the author’s philosophy appears in his remarks introducing “The Bernoulli numbers and some wonderful discoveries of Euler.” He derives the MacLaurin series for the cotangent via the generating function involving Bernoulli numbers, and also the partial fraction decomposition of the cotangent from Euler’s infinite product expansion for the sine. Comparing coefficients of the two expansions yields the beautiful formula

$$\sum_{i=1}^{\infty} \frac{1}{i^{2m}} = \frac{(-1)^{m+1} B_{2m} 2^{2m-1} \pi^{2m}}{(2m)!}.$$

In introducing these derivations, acknowledging that not every step is fully supported by the rigor available in a calculus class, Simmons says: “[The] mere fact that we are not able to seal every crack in the reasoning seems a flimsy excuse for denying students an opportunity to glimpse some of the wonders that can be found in this part of calculus.”

The division between the two parts of the book is not as great as their titles suggest. The *Brief Lives* are chock full of small mathematical gems and the history and development of concepts, often beautifully woven into the context of biography. An excellent example is Simmons’ treatment of Fermat. Along with the biography we find a comparison of the ideas of Descartes and Fermat on analytic geometry, and the details of Fermat’s method of finding tangents, complete with a quote from Newton crediting Fermat for his own ideas. There is also reference to a ‘nugget’ in the second part of the book giving Fermat’s ingenious method for calculating $\int_0^b x^n dx$ by using a geometric, rather than arithmetic, partition of the interval. We find ties to modern science, beginning with Fermat’s proof of Snell’s refraction law from his principle of least time, and discussion of how this led to the calculus of variations and Hamilton’s principle of least action. And finally, a comparison of the discoveries of Fermat in calculus with those of Newton and Leibniz, and a description of several of his important results in number theory.

The biographies are beautifully written and make engaging reading. The style is often captivating and witty, occasionally irreverent, and sometimes wryly humorous, as in the remark that Euler’s “personal life was as placid and uneventful as is possible for a man with 13 children” or that Cauchy was “a great mathematician who happened also to be a sincere and narrow-minded bigot.” Many of the

biographies are prose gems worth reading out loud in class, as I often do with Euler:

“... He was perhaps the most prolific author of all time in any field. From 1727 to 1783 his writings poured out in a seemingly endless flood, constantly adding to every known branch of pure and applied mathematics, and also to many that were not known until he created them. He averaged about 800 printed pages a year throughout his long life, and yet he almost always had something worthwhile to say and never seems long-winded. ... Euler evidently wrote mathematics with the ease and fluency of a skilled speaker discoursing on subjects with which he is intimately familiar. His writings are models of relaxed clarity. He never condensed, and he reveled in the rich abundance of his ideas and the vast scope of his interests. The French physicist Arago, in speaking of Euler’s incomparable mathematical facility, remarked that ‘He calculated without apparent effort, as men breathe, or as eagles sustain themselves in the wind.’ ... He was also a man of broad culture, well versed in classical languages and literatures (he knew the *Aeneid* by heart), many modern languages, physiology, medicine, botany, geography, and the entire body of physical science as it was known in his time. However, he had little talent for metaphysics or disputation, and came out second best in many good-natured verbal encounters with Voltaire at the court of Frederick the Great.”

The thirty-three mathematical biographies have a vast reach; from Thales in the 7th century B.C. to Weierstraß near the turn of our own waning century. They are not restricted to calculus, but include for example the development of the concept of mathematical proof from the Greek tradition of criticism and skepticism. In fact the biographies touch in a meaningful way on virtually every aspect of mathematics up into the nineteenth century. For instance, we learn about the astonishing discovery by Heiberg in 1906 of *The Method* by Archimedes, and the train of ideas leading from Apollonius’ *Conics* to Kepler’s laws of planetary motion and thence to Newton’s theory of gravitation and laws of motion. We are treated to a derivation of the planar forms for conic sections arising from sectioning a cone, and why these are called ‘parabola, hyperbola, ellipse’; and also to the classification of Pythagorean triples. We learn Cavalieri’s Principle for computing volumes by his ‘geometry of indivisibles,’ and are invited to try our own hand at this in a collection of problems (a few sections of the book have good problem sets; more would be a welcome addition). Simmons sometimes expresses strong opinions, for example “it was Cavalieri the disciple and not Galileo the master who first published the correct parabolic law of projectile motion, which most scientists assume is due to Galileo. In fact, Galileo was not much of a mathematician, and managed to confuse himself about the geometry.”

The cycloid is a recurring theme scattered throughout the book. When reading about Torricelli we learn that Galileo incorrectly concluded that the area under one arch was not quite three times the area of the generating circle, but that his student Torricelli proved it was exactly that; and the next major step in its study was Christopher Wren’s 1658 demonstration that its length is four times the diameter of the circle. We also learn that in 1645 Torricelli performed the very first calculation of the length of a noncircular curve, namely the equiangular spiral, and how difficult this was to believe: Descartes had asserted it was impossible,

“Geometry should not include lines that are like strings, in that they are sometimes straight and sometimes curved, since the ratios between straight and curved lines are not known, and I believe cannot be discovered by human minds,” while Galileo responded, “Who is so blind as not to see that, if there are two equal straight lines, one of which is then bent into a curve, that curve will be equal to the straight line?”

The story of the cycloid continues, not only through the biographies of Huygens, Newton, and the Bernoulli brothers, but also in four nuggets in the second part on the catenary, brachistochrone, tautochrone, and evolutes and involutes of the cycloid. For each a full and delightful mathematical treatment is presented, and the prominence of this 17th century theme comes through clearly, for instance in Jean Bernoulli’s “... you will be petrified with astonishment when I say that this very same cycloid, the tautochrone of Huygens, is also the brachistochrone we are seeking,” and his reaction to Newton’s anonymous solution: “I recognize the lion by his print.” This cycloid theme is so captivating that I wish it were graced with a unified exposition.

The mathematical biographies of Newton and Leibniz are strong, fascinating, and awe inspiring. I have the sense that Simmons’ own love and admiration is perhaps greatest for the 17th century, stretching here from Descartes and Kepler to Newton, Leibniz, and the Bernoullis, as evidenced by his quote from Alfred North Whitehead:

“A brief, and sufficiently accurate, description of the intellectual life of the European races during the succeeding two centuries and a quarter up to our own times is that they have been living upon the accumulated capital of ideas provided for them by the genius of the seventeenth century. The men of this epoch inherited a ferment of ideas attendant upon the historical revolt of the sixteenth century, and they bequeathed formed systems of thought touching every aspect of human life. It is the one century which consistently, and throughout the whole range of human activities, provided intellectual genius adequate for the greatness of its occasions.”

Continuing, the mathematical biography of Euler is equally awe inspiring, but by the mid-18th century, Simmons’ biographies become on average somewhat weaker, more personal biography and less mathematical biography, and I sense that this is not just because there are fewer giants and more advanced mathematics to explain. There are exceptions, with lovely, fascinating writings on both the lives and work of Gauß, Riemann, and Weierstraß, and to a lesser extent Liouville and Chebyshev, but weakness in those on Lagrange, Laplace, Fourier, Cauchy, Abel, and Dirichlet. Simmons gives a beautiful exposition of Riemannian geometry, and Riemann’s work on the prime number theorem via the zeta function. The wonderful Weierstraß biography, newly written for this book, has lucid descriptions of how elliptic functions arose as natural generalizations of trigonometric functions, and of spacefilling curves, and a fascinating footnote on the work of Bolzano. Thus this part of the book ends with a bang; one can hardly imagine a more enticing historical introduction to the ideas of higher mathematics for a student of calculus.

The author’s view of the importance of making the contribution of these mathematicians known and appreciated is indicated by his quote from Arthur

Koestler:

“In the index to the six hundred odd pages of Arnold Toynbee’s *A Study of History*, abridged version, the names of Copernicus, Galileo, Descartes and Newton do not occur...yet their cosmic quest destroyed the mediaeval vision of an immutable social order in a walled-in universe and transformed the European landscape, society, culture, habits and general outlook, as thoroughly as if a new species had arisen on this planet.”

The biographies also make a wealth of specific humanistic, psychological, social, and scientific connections regarding the development of Western civilization, with tantalizing footnotes and suggestions for further reading, and marvelous quotations throughout. For instance, regarding discovery, “It is often seen in the history of ideas that one generation’s disaster is an opportunity for the next,” or Linus Pauling’s comment that the way to have many good ideas is “Have lots of ideas and throw away the bad ones. You aren’t going to have good ideas unless you have lots of ideas and some sort of principle of selection.” Again, Simmons often expresses strong opinions: He first quotes the response of English political philosopher Thomas Hobbes to Cavalieri’s determination that a solid (the hyperbolic solid of revolution) can have infinite extent but finite volume, as “To understand this for sense, it is not required that a man should be a geometrician or a logician, but that he should be mad,” and then goes on to say “Hobbes had the curious belief that mathematical theorems can be attacked by ridicule and invective as if they were obnoxious planks in an opponent’s political platform.” However, some of the quotations are unfortunately stated without reference, and in general I wish that some of the claims, as profound and provocative as they are, were better documented, lest the reader wonder whether the desire for a captivating story sometimes takes over.

Some of the Memorable Mathematics pieces in the second part have already been mentioned above, a few more I will mention now, leaving yet others for the reader to discover when she/he reads the book. All are eminently accessible to calculus students, and well worth presenting or having students work on as projects.

- Several proofs of the Pythagorean Theorem, including two by the Indian mathematician Bhaskara, are followed by beautiful proofs of Heron’s formula for the area of a triangle in terms of its side lengths, and of Brahmagupta’s generalization to quadrilaterals (great for a precalculus class?) There seems to be no other work by non-Western mathematicians in the book, however.
- A beautiful description of Archimedes’ quadrature of a segment of a parabola, finding it is $\frac{4}{3}$ the area of a certain inscribed triangle (perfect for the second week of a ‘lean and lively’ calculus course?).
- The earliest, alluring, precise determination of the area of a region bounded by curves (no, it’s not a circle, and I won’t give it away).
- An exposition of Archimedes’ *Method*, the equilibrium of moments in balance he described in his letter to Eratosthenes for finding the volume of a sphere. The earliest appearance of the basic idea of integral calculus, this is quite suitable for reading and discussion with first semester students. Archimedes himself said “I am persuaded that this method will be of no little service to mathematics. For I foresee that once it is understood and established, it will

be used to discover other theorems which have not yet occurred to me, by other mathematicians, now living or yet unborn.” How ironic that this treatise was lost until the year 1906!

- Several gems deriving particular infinite series and products, such as Wallis’ infinite product for π , Leibniz’s ingenious derivation of the series for $\pi/4$ which bears his name using integration by parts, and three delightful derivations of Euler’s summation of $\Sigma(1/n^2)$, including the master’s own method based on factoring $(\sin x/x)$ into an infinite product.
- The distribution of primes.
- The irrationality of π , and of rational powers of e , Liouville’s Theorem on approximating transcendentals with rationals, and Hilbert’s version of Hermite’s proof that e is transcendental.
- The product form of the zeta function, and a proof that the series of reciprocals of the primes diverges.
- Derivations of Kepler’s three laws (including some nice problems).

The gems have a sumptuous finale, beginning with the extension of the complex numbers to the quaternions, discussion of more general abstract “linear algebras,” nonassociative algebras, and then an impressive application of the quaternions to prove Lagrange’s theorem that every natural number can be expressed as the sum of four squares. (This is an example of several spots where more historical perspective could enrich the gems, and more cross references to the *Brief Lives* would be illuminating.)

Simmons ends by enticing readers in further directions with the Armenian proverb that “mathematics, love, and home remodeling are the principal human enterprises in which one thing leads to another.”

The somewhat artificial division between the “*Brief Lives*” and the “*Memorable Mathematics*” could be bridged in large part by the incorporation of original sources. Using the words and methods of its creators would unify the mathematics with its human and developmental context, and bring it even more alive. It is hard to overestimate the positive impact that reading original sources can have on learning mathematics.

Every mathematician will learn something new and interesting from Simmons’ book. It shows calculus (and more) as a lively, powerful, and challenging human struggle. This is what both we and our students need to feel, and Simmons reminds us of its wonders from a historical perspective. Students who work with these ideas will gain motivation and awe we otherwise fail to provide.

When my niece started college at a (very large) premier west coast university, where her calculus class numbers several hundred students, I gave her a copy of *Calculus Gems*, in a desperate attempt to counteract all that she will encounter there militating to convince her that calculus is a class well worth dropping, not treasuring. As with everything about an eighteen year old, only time will tell.

Department of Mathematical Sciences
New Mexico State University
Las Cruces NM 88003
davidp@nmsu.edu

TELEGRAPHIC REVIEWS

Edited by **Arnold Ostebee and Paul Zorn**

with the assistance of the Mathematics Departments of
Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T : Textbook	P : Professional Reading	1-4 : Semester
C : Computer Software	L : Undergraduate Library	** : Special Emphasis
S : Supplementary Reading	13 : Grade Level	?? : Questionable

Readers are advised that price information is subject to change. Selected books and software packages receive a second, more extensive review in the *Monthly*.

Books and software submitted for review should be sent to *Book Reviews Editor*, *American Mathematical Monthly*, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098.

General, P, L*.** *Handbook of Writing for the Mathematical Sciences*. Nicholas J. Higham. SIAM, 1993, xii + 241 pp, \$21.50 (P). [ISBN 0-89871-314-5] The mathematician's Strunk and White. With chapters such as English Usage, Writing a Paper, Revising a Draft, Writing a Talk, and Computer Aids for Writing and Research, this handbook belongs on every mathematician's desk. A pleasure just to thumb through and spot read. BC

General, P. *How To Teach Mathematics: A Personal Perspective*. Steven G. Krantz. AMS, 1993, x + 76 pp, \$21 (P). [ISBN 0-8218-0197-X] Practical advice about teaching. Few will agree with every recommendation, but there is much of value here. AO

General, S(13). *Conquering Math Anxiety: A Self-Help Workbook*. Cynthia A. Arem. Brooks/Cole, 1993, xvii + 158 pp, \$9.50 (P). [ISBN 0-534-18876-1] Hints and exercises for controlling anxiety, developing positive thinking, visualizing success, recognizing and applying learning styles, learning study skills, conquering test anxiety. Despite the title, this is a general approach; main references to mathematics are chapter and section titles. MW

Mathematics Appreciation, T(13: 1), S, L*. *Geometry in Nature*. Vagn Lundsgaard Hansen. Transl: Tom Artin. AK Peters, 1993, xiii + 238 pp, \$29.95. [ISBN 1-56881-005-9] A broad, accessible, readable overview. Topics include: classification of surfaces, Poincaré's conjecture, catastrophe theory, ancient astronomy, relativity, Minkowski space-time, fiber bundles, gauge theory, string theory. DP

Recreational Mathematics, S(13), L. *Mathematics: How To Look Like a Genius Without Really Trying*. Arthur Benjamin, Michael Brant Shermer. Lowell House (2029 Century Park E., Suite 3290, Los Angeles, CA 90067), 1993, xx + 218 pp, \$22.95. [ISBN 0-929923-54-5] Dedicated to the proposition that learning mental arithmetic can hook students on real mathematics, this book explains tricks Arthur Benjamin uses when performing as a "lightning calculator." This reviewer was charmed by the fun-filled anecdotes, clever tricks, and wonderful quotes such as this one, from Samuel Johnson: "Arithmetical inquiries give entertainment in solitude by the practice, and reputation in public by the effect." AWR

Precalculus, T(13: 1). *Algebra with Trigonometry for College Students, Third Edition*. Charles P. McKeague. Saunders College, 1993, xix + 827 pp, \$46.75. [ISBN 0-03-096561-6] Many exercises in each section; extensive review material after each chapter. Useful chapter summaries, "common mistakes" sections. Clearly worked examples, with helpful hints, common sense suggestions. (*First Edition*, TR, December 1988.) DS

Precalculus, C. *GrafEq V1.13 for the Apple Macintosh*. Pedagogy Software (4446 Lazelle Ave., Terrace, BC, Canada V8G 1R8), 1991, 21 pp, (P), \$200 site license. A simple, inexpensive, easy-to-use plotting program with pre-calculus and elementary calculus applications. Plots equations—not just functions; includes convenient zooming, tracking, deriva-

tive, and other features. Manual includes sample teaching materials. PZ

Education, P*. *Integrating Research on the Graphical Representation of Functions.* Eds: Thomas A. Romberg, Elizabeth Fennema, Thomas P. Carpenter. Stud. in Math. Thinking & Learning. Lawrence Erlbaum Assoc, 1993, xi + 350 pp, \$79.95. [ISBN 0-8058-1134-6] 11 articles summarize current knowledge from teaching, learning, curriculum, and assessment perspectives. AO

Education, P. *Open-ended Questioning: A Handbook for Educators.* Robin Lee Harris Freedman. A-W, 1994, v + 81 pp, \$12.96 (P). [ISBN 0-201-81958-9] Advice on writing and using open-ended questions for assessment. Some examples from mathematics. AO

Education, P. *Authentic Assessment: A Handbook for Educators.* Diane Hart. A-W, 1994, vii + 120 pp, \$12.96 (P). [ISBN 0-201-81864-7] Surveys alternatives to standardized tests for assessing learning. Gives practical tips for use of authentic assessment techniques. AO

Education, P, L. *Assessment in the Mathematics Classroom: 1993 Yearbook.* Norman L. Webb, Arthur F. Coxford. NCTM, 1993, viii + 248 pp, \$20. [ISBN 0-87353-352-6] 27 papers on general assessment themes; assessment techniques and practices in grades K-4, 5-8, and 9-12; issues in assessment. AO

Education, P, L. *Partnerships in Maths: Parents and Schools: The IMPACT Project.* Eds: Ruth Merttens, Jeff Vass. Falmer Pr, 1993, vi + 255 pp, \$35 (P); \$90. [ISBN 0-75070-155-2; 0-75070-154-4] The IMPACT project made parents partners in mathematics education. 22 papers, written by parents, teachers, supervisory personnel, and researchers, offer perspectives on the project and guidance for those interested in similar efforts. AO

Education, P. *American Perspectives on the Seventh International Congress on Mathematical Education.* Ed: John A. Dossey. NCTM, 1993, vi + 73 pp, \$10 (P). [ISBN 0-87353-360-7] 38 papers by American educators report on sessions, experiences, and observations. AO

Education, P. *Mathematics Assessment: Alternative Approaches.* Therese Kuhs. NCTM, 1992, \$55 video/viewer's guide set; A Viewer's Guide, iv + 20 pp, \$4.50 (P). [ISBN 0-87353-355-0] Dramatizations illustrate several approaches to assessment. Viewer's Guide summarizes each video segment, suggests follow-up activities. Stimulating, provocative. AO

Education, S(15-18). *Talking Points in Mathematics.* Anita Straker. Cambridge Univ Pr, 1993, 124 pp, \$15.95 (P). [ISBN 0-521-44758-

5] Ideas for mathematical "thought experiments" in elementary and middle school. Activities and teaching suggestions to encourage mathematical discussion, imagery, and mental arithmetic. Organized by mathematical themes and British school levels. MW

History, P. *Bertrand Russell and the Origins of the Set-theoretic 'Paradoxes'.* Alejandro R. Garciadiego. Birkhäuser, 1992, xxix + 264 pp, \$77.50. [ISBN 0-8176-2669-7] "Reconstructs and reinterprets the role of Russell in the origins of the set-theoretic 'paradoxes'." Carefully documented scholarly work; extensive bibliography. BH

Logic, T(17), P, L. *Model Theory.* Wilfrid Hodges. Ency. of Math. & Its Applic., V. 42. Cambridge Univ Pr, 1993, xiii + 772 pp, \$99.95. [ISBN 0-521-30442-3] Nice encyclopedic introduction to model theory; chapters loosely organized around theme of model-theoretic constructions. RM

Logic, T(16-18: 1, 2), L. *Logic and Prolog.* Richard Spencer-Smith. Harvester Wheatsheaf (Div. of Simon & Schuster), 1991, viii + 392 pp. [ISBN 0-7450-1022-9] Stimulating interplay between classical logic and Prolog. Logic is first seen as a representation system for knowledge found in natural language. Then Prolog as a programming language for processing knowledge is discussed. In the second half, logic as a system of deduction, is followed by considering Prolog as a computer system for achieving similar results. RJA

Combinatorics, T(17-18: 2), P, L. *Algebraic Combinatorics.* C.D. Godsil. Chapman & Hall, 1993, xv + 362 pp, \$54.95. [ISBN 0-412-04131-6] Explores interactions between algebra and combinatorics. Develops theory of characteristic and matching polynomials of a graph, connection to orthogonal polynomials. Also treats connections among orthogonal polynomials and other combinatorial objects (e.g., walks), distance regular graphs, association schemes, polynomial spaces. LC

Linear Algebra, T(14-16: 1), L. *Linear Algebra: An Introduction to Abstract Mathematics.* Robert J. Valenza. Undergrad. Texts in Math. Springer-Verlag, 1993, xviii + 237 pp, \$39. [ISBN 0-387-94099-5] Distinctly different from the usual computational approach, this treatment leans heavily toward axiomatic structural development. Matrices don't appear until Chapter 5. Non-standard topics at this level include dual spaces, the Cayley-Hamilton theorem, and the Jordan normal form. Supplementary topics include quadratic forms, and cat-

egories and functors. Lean, concise treatment; exercises often illustrate or extend theory. JS

Group Theory, T(18: 1, 2), P. *Nilpotent Groups and their Automorphisms*. Evgenii I. Khukhro. Expos. in Math., V. 8. Walter de Gruyter, 1993, xiii + 252 pp, DM 158. [ISBN 3-11-013672-4] First half treats basic linear and combinatorial methods in theory of nilpotent groups. Second half details recent results. DP

Group Theory, T*(16-17: 1, 2), L. *Representations and Characters of Groups*. Gordon James, Martiñ Liebeck. Cambridge Univ Pr, 1993, x + 419 pp, \$69.95; \$29.95 (P). [ISBN 0-521-44024-6; 0-521-44590-6] Nice introduction to group representations, based on modules but stressing characters. Useful worked examples (including character tables for many groups), good exercises (many with solutions) make text accessible to undergraduates, chemists, physicists, etc. Appealing application to molecular vibration. RM

Algebra, T(17: 1, 2), L. *Algebra*. T.T. Moh. Ser. on Univ. Math., V. 5. World Scientific, 1992, viii + 350 pp, \$68; \$32 (P). [ISBN 981-02-1195-3; 981-02-1196-1] Provides "quick access to the basic knowledge of algebra," while showing diverse applications. Many examples; exercises stress intuitive approach. DP

Algebra, P. *The Basic Theory of Power Series*. Jesús M. Ruiz. Adv. Lect. in Math. Friedr Vieweg & Sohn, 1993, ix + 134 pp, \$17 (P). [ISBN 3-528-06525-7] Theorems and techniques of power series for analytic and algebraic geometry, commutative algebra. Includes Rückert's complex nullstellensatz, Risler's real nullstellensatz, Tougeron's implicit function theorem, Artin's approximation theorem. BH

Calculus, T(13: 2). *Calculus with Applications, Second Edition*. Raymond F. Coughlin, David E. Zitarelli. Saunders College, 1993, xxii + 722 pp, \$58.75. [ISBN 0-03-055757-7] Changes (*First Edition*, TR, August-September 1989) include exercises divided into three sets: standard problems, "referenced exercises" (new and updated applications problems), "cumulative exercises" (solutions require ideas from preceding sections). Beginner's appendix for graphics calculators. Graphics calculator explorations in each chapter. KB

Calculus, T(13). *Applied Technical Mathematics*. Merwin J. Lyng, L.J. Meconi, Earl J. Zwick. Wm C Brown, 1992, xvi + 857 pp. [ISBN 0-697-08543-0]; *Applied Technical Mathematics with Calculus*, xvi + 1142 pp. [ISBN 0-697-05970-7] Basic algebra, trigonometry, sequences and series, analytic geometry, statistics; calculus in the extended version. No-frills

approach stresses physical examples, not theory. Useful as a reference. BH

Calculus, T(13), P, L. *Calculus with Maple V*. John S. Devitt. Brooks/Cole, 1993, xxii + 502 pp, \$24.50 (P). [ISBN 0-534-16362-9] Standard topics and order of presentation, but Maple V replaces paper-and-pencil for computations. Exposition illustrates commands; style allows students to focus on concepts. End-of-chapter exercises; some go beyond what is reasonable to do "by hand." AO

Real Analysis, T(15-16: 1), P*, L.** *A Course of Pure Mathematics, Tenth Edition*. G.H. Hardy. Cambridge Univ Pr, 1992, xii + 509 pp, \$22.95 (P). [ISBN 0-521-09227-2] A reprint of the timeless classic, first published in 1908. Introduces analysis with the enthusiasm of "a missionary talking to cannibals." DP

Complex Analysis, T*(18), P*. *Composition Operators and Classical Function Theory*. Joel H. Shapiro. Universitext. Springer-Verlag, 1993, xvi + 223 pp, \$34 (P). [ISBN 0-387-94067-7] A holomorphic function ϕ that maps the unit disk U into itself defines the composition operator $C_\phi f = f \circ \phi$ for f holomorphic on U . Starting with Littlewood's subordination principle (that C_ϕ maps H^2 into itself), text explores "connections between function theoretic properties of ϕ and the behavior of C_ϕ on H^2 ." Rudin's *Real and Complex Analysis* is the only prerequisite. Chapter exercises; extensive bibliography. Wonderfully motivated exposition of some beautiful theory. BH

Differential Equations, T(18), P. *Singular Perturbation Methods for Ordinary Differential Equations*. Robert E. O'Malley, Jr. Appl. Math. Sci., V. 89. Springer-Verlag, 1991, viii + 225 pp. [ISBN 0-387-97556-X] Covers singularly perturbed initial value and boundary value problems, significant applications. Appendix on history. BL

Differential Equations, P. *Nonlinear Equations in the Applied Sciences*. Eds: W.F. Ames, C. Rogers. Math. in Sci. & Eng., V. 185. Academic Pr, 1992, ix + 475 pp. [ISBN 0-12-056752-0] 11 papers on symmetry in nonlinear mechanics, waves, integrable systems, symmetric chaos, etc. BL

Differential Equations, S(18), P, L. *Solving Ordinary Differential Equations I: Nonstiff Problems, Second Revised Edition*. E. Hairer, S.P. Nørsett, G. Wanner. Ser. in Comput. Math., V. 8. Springer-Verlag, 1993, xv + 528 pp, \$98. [ISBN 0-387-56670-8] New features: Hamiltonian systems, symplectic and parallel Runge-Kutta methods, dense output for RK methods, a new Dormand and Prince

method of order 8 with dense output. (*First Edition*, TR, June-July 1987.) SM

Partial Differential Equations, T(15-16: 1), L. *Boundary Value Problems and Partial Differential Equations.* Mayer Humi, William B. Miller. PWS-Kent, 1992, x + 342 pp, \$39. [ISBN 0-534-92880-3] For science and engineering students—intuitive, technique-oriented. Treats separation of variables, closed form representations, Fourier and Laplace transforms, numerical methods. BL

Partial Differential Equations, T(15: 2), L. *Partial Differential Equations and Boundary-Value Problems with Applications, Second Edition.* Mark A. Pinsky. McGraw-Hill, 1991, xxii + 461 pp. [ISBN 0-07-050128-9] Stresses solutions of physically motivated problems via separation of variables, Fourier series, and transforms. Also includes asymptotic analysis of integrals, numerical solutions, Green's functions, and approximate solutions. Appendix with Mathematica examples. (*First Edition*, TR, February 1985.) BL

Partial Differential Equations, P. *Pseudodifferential Operators and Nonlinear PDE.* Michael E. Taylor. Progress in Math., V. 100. Birkhäuser, 1991, 213 pp. [ISBN 0-8176-3595-5] Operators with nonsmooth symbols, paradifferential operators, nonlinear hyperbolic systems, propagation of singularities, nonlinear parabolic equations, elliptic boundary problems. BL

Dynamical Systems, P. *The Geometry of Hamiltonian Systems.* Ed: Tudor Ratiu. Math. Sci. Res. Inst., V. 22. Springer-Verlag, 1991, x + 527 pp. [ISBN 0-387-97608-6] 19 papers from 1989 MSRI workshop. BL

Dynamical Systems, T(18: 1, 2), P. *Elements of Topological Dynamics.* J. de Vries. Math. & Its Applic., V. 257. Kluwer Academic, 1993, xvi + 748 pp, \$235. [ISBN 0-7923-2287-8] Systematic exposition; aims at recent research. Assumes general topology, some Lebesgue integration theory and functional analysis. DP

Dynamical Systems, T(18), S, P, L. *Complex Dynamics.* Lennart Carleson, Theodore W. Gamelin. Universitext: Tracts in Math. Springer-Verlag, 1993, ix + 175 pp, \$29 (P). [ISBN 0-387-97942-5] Expanded notes from a UCLA course. Conformal and quasiconformal mappings, fixed points, rational iteration theory (Julia sets), classification of periodic components, critical points and expanding maps, quasiconformal surgery, local geometry of the Fatou set, and the Mandelbrot set. Assumes graduate real and complex analysis. KS

Numerical Analysis, S(16). *Numerical Meth-*

ods for Two-Point Boundary-Value Problems. Herbert B. Keller. Dover, 1992, 397 pp, \$9.95 (P). [ISBN 0-486-66925-4] A minor revision of the original 1968 Blaisdell text (TR, January 1969). DH

Numerical Analysis, T(13), L. *Numerical Mathematics—A Laboratory Approach.* Shlomo Breuer, Gideon Zwas. Cambridge Univ Pr, 1993, xiii + 267 pp, \$49.95. [ISBN 0-521-44040-8] Assumes neither calculus nor linear algebra. Topics include iterative processes, area approximations, linear systems, acceleration of convergence, interpolative approximation. AO

Numerical Analysis, T(15: 1), P. *Numerical Methods for Engineering Applications.* Edward R. Champion, Jr. Mech. Eng., V. 84. Marcel Dekker, 1993, xv + 442 pp, \$150. [ISBN 0-8247-9135-5] Covers the standard topics plus Fourier transforms. Includes application to partial differential equations, comments on some software packages. Chapters contain source code for each algorithm, examples. DH

Numerical Analysis, T(15-17: 1), L. *Solving Linear and Non-Linear Equations.* Chris Woodford. Math. & Its Applic. Ellis Horwood, 1992, 190 pp, \$60 (P). [ISBN 0-13-830415-7; 0-13-830423-8] Theoretical background and practical advice. A careful study of selected algorithms, not a comprehensive survey. AO

Functional Analysis, T(18: 1), S, P. *A Primer of Nonlinear Analysis.* Antonio Ambrosetti, Giovanni Prodi. Stud. in Adv. Math., V. 34. Cambridge Univ Pr, 1993, viii + 171 pp, \$44.95. [ISBN 0-521-37390-5] Differentiation in Banach spaces, including local and global inversion theorems; bifurcation theory applied to water waves, the restricted three-body problem, etc. Assumes solid grounding in linear functional analysis. Brief problem list. JS

Analysis, P. *Analysis and Geometry on Groups.* N. Th. Varopoulos, L. Saloff-Coste, T. Coulhon. Tracts in Math., V. 100. Cambridge Univ Pr, 1992, xii + 156 pp, \$44.95. [ISBN 0-521-35382-3] Advanced research material from courses given by Varopoulos at Université Paris VI, 1982-87. DP

Analysis, T(16-18), S, P, L. *Wavelets and Operators.* Yves Meyer. Transl: D.H. Salinger. Stud. in Adv. Math., V. 37. Cambridge Univ Pr, 1992, xv + 223 pp, \$49.95. [ISBN 0-521-42000-8] Surveys popular alternative to traditional Fourier analysis. Argues that Fourier and wavelet analysis are complementary. Assumes rudiments of Fourier analysis. A "must read" for anyone who wants to learn the subject. KS

Analysis, S(15-18), L. *Selected Problems in Real Analysis.* B.M. Makarov, et al. Transl.

of Math. Mono., V. 107. AMS, 1992, x + 370 pp, \$112. [ISBN 0-8218-4559-4] Problems (with solutions) for students who want to learn by working (mostly) hard problems. Arranged like Halmos's *A Hilbert Space Problem Book*. Chapters on sequences, functions, series, integrals, asymptotics, Lebesgue measure and integration, sequences of measurable functions, iterates of transformations. Based on a course at Leningrad University. KS

Analysis, T(18: 1, 2), P.** *The Logarithmic Integral: II*. Paul Koosis. Cambridge Univ Pr, 1992, xxvi + 574 pp, \$150. [ISBN 0-521-30907-7] *Volume II* of a comprehensive study of the logarithmic integral's wide-ranging roles in real and complex analysis. Like *Volume I* (TR, May 1989), an expository masterpiece: sophisticated but informal, serious but engaging, demanding but generous. PZ

Algebraic Geometry, S(18), P. *Clifford Numbers and Spinors*. Marcel Riesz. Eds: E. Folke Bolinder, Pertti Lounesto. Fund. Theories of Physics, V. 54. Kluwer Academic, 1993, ix + 241 pp, \$98.50. [ISBN 0-7923-2299-1] First section is transcription of author's 1957–1958 lectures on Clifford Algebras and Spinors; topics are Clifford numbers, rotations and reflections, canonical representations, representations of isometries. Second section is a supplement on Clifford algebras. Third section, by Lounesto, is a survey article on Clifford algebras. JS

Topology, T(16–17: 1, 2), L*.** *Classical Topology and Combinatorial Group Theory, Second Edition*. John Stillwell. Grad. Texts in Math., V. 72. Springer-Verlag, 1993, xi + 334 pp, \$49. [ISBN 0-387-97970-0] A wonderful textbook. Historical development stresses geometric aspects, focusing on connections with complex analysis, mechanics, group theory. Treats only dimensions ≤ 3 , with emphasis on fundamental group. New edition includes proof of the unsolvability of the word problem for groups, some of its consequences. (*First Edition*, TR, January 1982.) DP

Operations Research, S(16–17), C, L. *A Computer-Assisted Analysis System for Mathematical Programming Models and Solutions: A User's Guide for ANALYZE*. Harvey J. Greenberg. OR/Comp. Sci. Interface Ser. Kluwer Academic, 1993, xii + 264 pp, \$105, disk included. [ISBN 0-7923-9322-8] Many examples; no exercises. Assumes basics of LP formulation, solution, and post-optimality analysis. Can be used with MINOS, MPS III, MPSX, OB1, OSL, etc. SM

Mathematical Modeling, P. *Innovation in*

Maths Education by Modelling and Applications. Jan de Lange, et al. Math. & Its Applic. Ellis Horwood, 1993, xii + 392 pp, \$50. [ISBN 0-13-017351-7] 37 papers on philosophical issues, new topics and tools, case studies, curriculum and assessment. Many articles from the 1991 International Conference on the Teaching of Mathematics by Applications held in Noordwijkerhout, The Netherlands. AO

Probability, T(17–18: 2), S*. *Martingale Spaces and Inequalities*. Ruilin Long. Friedr Vieweg & Sohn, 1993, iv + 346 pp, \$64. [ISBN 3-528-08397-2] A systematic introduction; reflects developments of last 20 years. Topics: regular martingales and atomic decomposition; aspects related to BMO (Carleson measures, commutators, weights, martingale transforms, etc.); applications to analysis. KB

Probability, T(16–17: 2), S. *Probability and Its Applications for Engineers*. David H. Evans. Marcel Dekker, 1992, xiv + 634 pp, \$89.75. [ISBN 0-8247-8656-4] Six chapters on theory; six stand-alone chapters on applications to statistics, quality control, tolerancing, reliability, random processes, and decision trees. DH

Stochastic Processes, T(17–18: 1, 2), S, P, L. *Introduction to Multiple Time Series Analysis, Second Edition*. Helmut Lütkepohl. Springer-Verlag, 1993, xxi + 545 pp, \$49 (P). [ISBN 0-387-56940-5] Very minor changes from *First Edition* (TR, May 1992). KB

Elementary Statistics, S?(14). *The Proofs for a First Course in Statistics*. T.P. Hutchinson. Rumsby Scientific (POB Q355, Queen Victoria Bldg., Sydney, N.S.W. 2000, Australia), 1993, 10 pp, \$2.50 (P). Proofs of ten elementary results, such as $Var(\bar{X}) = Var(X)/n$ and the formulas for the coefficients of the least squares regression line. Requires some calculus. RSK

Statistical Methods, T(16–17: 2), S. *Time Series: Forecasting, Simulation, Applications*. Gareth Janacek, Louise Swift. Math. & Its Applic. Ellis Horwood, 1993, 331 pp. [ISBN 0-13-918459-7] Broad practical/theoretical introduction to time series analysis (mostly univariate). Stresses analysis and forecasting of state space models using Kalman filter for estimation. Topics: estimation, prediction, model selection and evaluation, ARMA models, Kalman filter algorithm, structural models, frequency domain ideas, spectral estimation, multiple series, inclusion of exogenous variables, transformations, missing values. KB

Computer Systems, P. *Understanding Japanese Information Processing*. Ken Lunde. O'Reilly & Assoc, 1993, xxxii + 435 pp, \$29.95 (P). [ISBN 1-56592-043-0]

Theory of Computation, T(17–18: 1), S, P. *Deduction and Declarative Programming.* Peter Padawitz. Tracts in Theoret. Comput. Sci., V. 28. Cambridge Univ Pr, 1992, vi + 279 pp, \$44.95. [ISBN 0-521-41723-6] Begins with basics of functional and logic programming, stressing specification, verification. Treats guards, generators, constructors, models, correctness, computing goal solutions, inductive and directed expansion, reduction, ground confluence. Some good examples. Last chapter treats EXPANDER, a proof support system for reasoning about data type specifications and declarative programs. RJ A

Theory of Computation, T(17–18: 1), S, P. *The Clausal Theory of Types.* D.A. Wolfram. Tracts in Theor. Comp. Sci. Cambridge Univ Pr, 1993, viii + 124 pp, \$39.95. [ISBN 0-521-39538-0] Begins by developing logic programming (LP) from automatic theorem proving. Uses simply typed λ -calculus to define syntax of higher-order logics and to study their properties. Uses higher-order equational unification to generalize the resolution principle. A higher-order LP language is derived and shown to be sound and complete. Extensive bibliography. RJ A

Computer Science, T(13: 1), L*. *The (New) Turing Omnibus: 66 Excursions in Computer Science.* A.K. Dewdney. Computer Science Pr, 1993, xvi + 455 pp, \$24.95. [ISBN 0-7167-8271-5] Revised and expanded; new chapters on disk operating systems, computer viruses, Newton's method, genetic algorithms, how neural nets learn, the Mandelbrot set. AO

Computer Science, P. *Specification and Proof in Real-Time CSP.* Jim Davies. Cambridge Univ Pr, 1993, xvii + 180 pp, \$49.95. [ISBN 0-521-45055-1] Extends Hoare's theory of Communicating Sequential Processes to timed denotational semantics for the specification of real-time systems. RM

Applications (Biological Science), T(15–17: 1), L. *Mathematical Modeling in the Life Sciences.* Paul Doucet, Peter B. Sloep. Math. & Its Applic. Ellis Horwood, 1992, xiv + 490 pp, \$45 (P). [ISBN 0-13-562018-X; 0-13-562000-7] Aims to present "what a life scientist should know about modeling." Differential equations and probability theory are the primary mathematical tools. AO

Applications (Economics), T(16–17: 2), S. *The Medieval Village Economy: A Study of the Pareto Mapping in General Equilibrium Models.* Robert M. Townsend. Frontiers of Econ. Res. Princeton Univ Pr, 1993, xxi + 145 pp, \$35. [ISBN 0-691-04270-5] Path-breaking contributions to contract theory and

general equilibrium analysis. Explains medieval village economy by combining theory of general equilibrium with notion that allocations and institutions of an economy might be Pareto optimal. Standard models are reinterpreted, applied, and extended. Draws from historical observations; characterizes solutions analytically and numerically. KB

Applications (Information Theory), P. *Situation Theory and Its Applications, Volume 2.* Eds: Jon Barwise, et al. Lect. Notes, No. 26. CSLI (Ventura Hall, Stanford U., Stanford, CA 94305), 1991, xiii + 637 pp, \$26.95 (P). [ISBN 0-937073-70-9] Proceedings of 1990 conference in Scotland. Papers treat issues in situation theory; applications of theory to logic, computer science, philosophy; applications to natural language. RJ A

Applications (Physical Science), S(18), L. *Lecture Notes in Control and Information Sciences-186: Systems Representation of Global Climate Change Models: Foundation for a Systems Science Approach.* N. Sreenath. Springer-Verlag, 1993, xxvi + 258 pp, \$79 (P). [ISBN 0-387-19824-5] Models global climate change using systems theory. Systems theory is used to model global climate change. An intricate block diagram on a fold-out page describes the model graphically. Stresses incorporation of human activities into the model. SM

Applications (Physics), P. *Introduction to Quantum Groups.* George Lusztig. Progress in Math., V. 110. Birkhäuser, 1993, xii + 341 pp, \$49.50. [ISBN 0-8176-3712-5]

Applications (Physics), P. *Mathematical Scattering Theory: General Theory.* F.R. Yafaev. Transl. of Math. Mono., V. 105. AMS, 1992, x + 341 pp, \$216. [ISBN 0-8218-4558-6]

Applications (Physics), P, L. *Selected Works of Yakov Borisovich Zeldovich, Volume II: Particles, Nuclei, and the Universe.* Eds: J.P. Ostriker, G.I. Barenblatt, R.A. Sunyaev. Transl: A. Granik, E. Jackson. Princeton Univ Pr, 1993, xiv + 644 pp, \$59.50. [ISBN 0-691-08742-3] 83 papers on nuclear and particle physics, astrophysics and cosmology, and the history of physics. BC

Reviewers

RJA: Richard J. Allen, St. Olaf; KB: Karla Ballman, Macalester; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; BH: Bruce Hanson, St. Olaf; DH: Deanna Haunsperger, St. Olaf; RSK: Richard S. Kleber, St. Olaf; BL: Brian Loe, Carleton; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; AO: Arnold Ostebee, St. Olaf; DP: David Peifer, St. Olaf; AWR: A. Wayne Roberts, Macalester; KS: Karen Saxe, Macalester; JS: John Schue, Macalester; DS: Dan Schwalbe, Macalester; MW: Martha Wallace, St. Olaf; PZ: Paul Zorn, St. Olaf.

Symbolic Computation in Undergraduate Mathematics Education

Zaven Karian, Editor

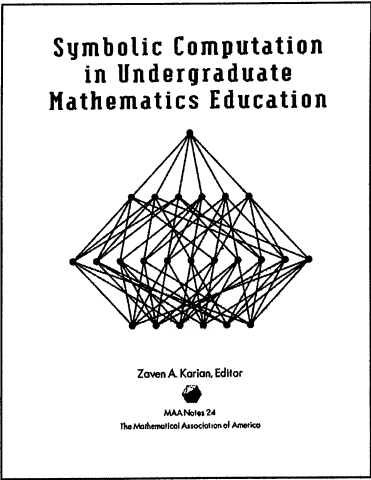
If you are considering putting a symbolic computing system into your curriculum, this is one publication you should have.
—Mathematics Teacher

This well-written book should be helpful to anyone using symbolic computation as an aid in teaching undergraduates—The book provides a number of examples for presenting probability and statistics in a way that removes the tedium and emphasizes the underlying ideas.
—AAAS, Science Book and Films

If you have any plans to integrate symbolic computing into your program, read and study this book first. Your students will thank you for it.
—AMATYC Review

This volume brings together many of the facets associated with the pedagogic uses of symbolic computation.

Part I consists of articles that deal with general issues of learning mathematics and the role of symbolic computation in that process. The articles in Part II describe the use of symbolic computation in teaching calculus. Some of the areas covered are the use of symbolic computation in a laboratory calculus course, the uses of Derive in the instruction of calculus, antidifferentiation and the



definite integral, and the experiences and reflections of teachers who have used symbolic computation in calculus instruction.

Part III consists of papers on sophomore-level courses on linear algebra and differential equations. The articles in Part IV describe what can be done in using symbolic computation in teaching combinatorics, probability and statistics courses. The articles and references in Part V will help you get started in using some of these ideas at your own institution.

200 pp., 1992, Paperbound
ISBN 0-88385-082-6
List: \$22.00
Catalog Number NTE-24

ORDER FROM:
The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 2003
1-800-331-1622 Fax (202) 265-2384

Membership Code	Qty.	Catalog Number	Price
_____	_____	_____	_____
Name _____	_____		
Address _____	_____		
City _____	Total \$ _____		
State _____ Zip Code _____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
	Credit Card No. _____		
	Signature _____ Exp. Date _____		

Proofs Without Words

Exercises in Visual Thinking

Roger B. Nelsen

Just what are “proofs without words?” First of all, most mathematicians would agree that they certainly are not “proofs” in the formal sense. Indeed, the question does not have a simple answer. Proofs without words are generally pictures or diagrams that help the reader see *why* a particular mathematical statement may be true, and *how* one could begin to go about proving it. While in some proofs without words an equation or two may appear to help guide that process, the emphasis is clearly on providing *visual* clues to stimulate mathematical thought. Proofs without words bear witness to the observation that often in the English language to *see* means to *understand*, as in “to see the point of an argument.”

Proofs without words have a long history. In this collection you will find modern renditions of proofs from ancient China, classical Greece, twelfth-century India—even one based on a published proof by a former President of the United States! However, most of the proofs are more recent creations, and many are taken from the pages of MAA journals.

The proofs in this collection are arranged by topic into six chapters: Geometry and Algebra; Trigo-

nometry, Calculus and Analytic Geometry; Inequalities; Integer Sums; Sequences and Series; and Miscellaneous. Teachers will find that many of the proofs in this collection are well suited for classroom discussion and for helping students to think visually in mathematics.

The readers of this collection will find enjoyment in discovering or rediscovering some elegant visual demonstrations of certain mathematical ideas that teachers will want to share with their students. Readers may even be encouraged to create new “proofs without words.”

160 pp., Paperbound, 1993

ISBN 0-88385-700-6

List: \$27.50 MAA Member: \$22.00

Catalog Number PWW

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty.	Catalog Number	Price
------	----------------	-------

_____	_____	_____
_____	_____	_____

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____



**No matter how you
express it, it still means
DERIVE® is half price.**

$$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} \quad \lim_{x \rightarrow 0} \frac{x}{\sin(2x)} \quad \frac{1}{2}$$

50%

$$\sum_{n=1}^{\infty} \frac{1}{2^{n+1}}$$

0.5

$$\int_0^1 x \, dx$$

DERIVE ➡

The *DERIVE A Mathematical Assistant* program lets you express yourself symbolically, numerically and graphically, from algebra through calculus, with vectors and matrices too—all displayed with accepted math notation, or 2D and 3D plotting. *DERIVE* is also easy to use and easy to read, thanks to a friendly, menu-driven interface and split or

overlay windows that can display both algebra and plotting simultaneously. Better still, *DERIVE* has been praised for the accuracy and exactness of its solutions. But, best of all the suggested retail price is now only \$125. Which means *DERIVE* is now half price, no matter how you express it.

System requirements

DERIVE: MS-DOS 2.1 or later, 512K RAM, and one 3½" disk drive. Suggested retail price now **\$125 (Half off!)**.

DERIVE ROM card: Hewlett Packard 95LX & 100LX Palmtop, or other PC compatible ROM card computer. Suggested retail price now **\$125!**

DERIVE XM (eXtended Memory): 386 or 486 PC compatible with at least 2MB of *extended* memory. Suggested list price now \$250!

DERIVE is a registered trademark of Soft Warehouse, Inc.

 **Soft Warehouse**
HONOLULU • HAWAII

Soft Warehouse, Inc. • 3660 Waiialae Ave.
Ste. 304 • Honolulu, HI, USA 96816-3236
Ph: (808) 734-5801 • Fax: (808) 735-1105

Game Theory and Strategy

Philip D. Straffin, Jr.



This valuable addition to the New Mathematical Library series pays careful attention to applications of game theory in a wide variety of disciplines. The applications are treated in considerable depth. The book assumes only high school algebra, yet gently builds to mathematical thinking of some sophistication. **Game Theory and Strategy** might serve as an introduction to both axiomatic mathematical thinking and the fundamental process of mathematical modelling. It gives insight into both the nature of pure mathematics, and the way in which mathematics can be applied to real problems.

Since its creation by John von Neumann and Oskar Morgenstern in 1944, game theory has contributed new insights to business, politics, economics, social psychology, philosophy, and evolutionary biology. In this book, the fundamental ideas of game theory share the stage with applications of the theory. How might strategic business decisions depend on information about a rival company, and how much would such information be worth? When is it advantageous to vote for a candidate who is not your favorite? What are the optimal strategies for teams in the football draft, and what paradoxes can result from following

those strategies? What is a fair way to share the costs of a development project? What can we learn about the problem of "free will" by imagining playing a game with an omnipotent Being? How might natural selection lead to altruistic behavior in animal species? Game theory gives insight into all of these questions.

The book includes many exercises, with answers, which allow the reader to try out calculations, and explore alternative formulations of game-theoretic ideas.

200 pp., 1993, Paperbound

ISBN 0-88385-637-9

List: \$27.50 MAA Member: \$22.00

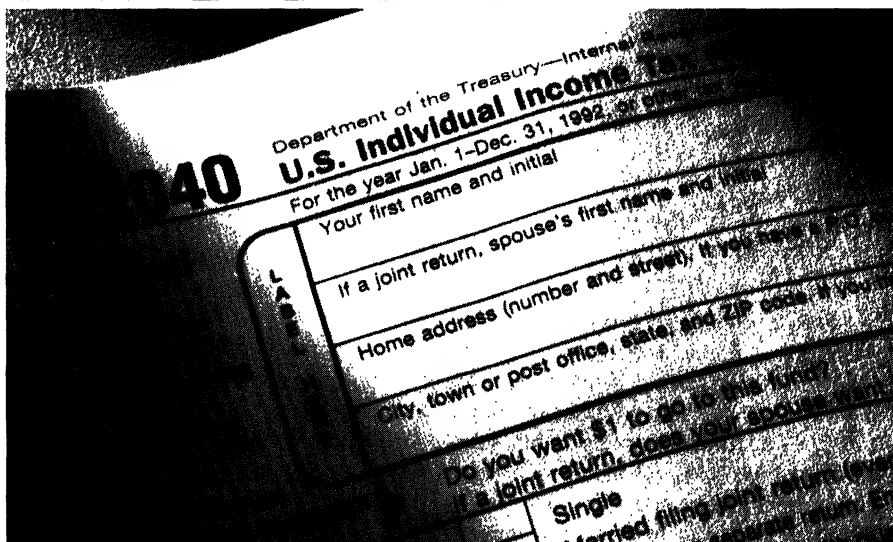
Catalog Number NML-36

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
(202) 387-5200 Fax (800) 331-1622

Membership Code -----	Qty.	Catalog Number	Price
Name _____	_____	_____	_____
Address _____	_____	_____	Total \$ _____
City _____	_____	_____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
State ____ Zip Code _____	_____	_____	Credit Card No. _____
	_____	_____	Signature _____
	_____	_____	Exp. Date _____

PRINCIPLES of SOUND RETIREMENT INVESTING



UNFORTUNATELY, THIS IS WHERE PEOPLE ARE PUTTING TOO MANY RETIREMENT DOLLARS.

Every year, a lot of people make a huge mistake on their taxes. They don't take advantage of tax deferral and wind up sending Uncle Sam money they could be saving for retirement.

Fortunately, that's a mistake you can easily avoid with TIAA-CREF SRAs. SRAs not only ease your current tax-bite, they offer a remarkably easy way to build retirement income—especially for the “extras” that your regular pension and Social Security benefits may not cover. Because your contributions are made in before-tax dollars, you pay less taxes now. And since all earnings on your SRA are tax-deferred as well, the

money you don't send to Washington works even harder for you. Down the road, that can make a dramatic difference in your quality of life.

What else makes SRAs so special? A range of allocation choices—from the guaranteed security of TIAA to the diversified investment accounts of CREF's variable annuity—all backed by the nation's number one retirement system.

Why write off the chance for a more rewarding retirement? Call today and learn more about how TIAA-CREF SRAs can help you enjoy many happy returns.

Benefit now from tax deferral. Call our SRA hotline 1 800-842-2733, ext. 8016.

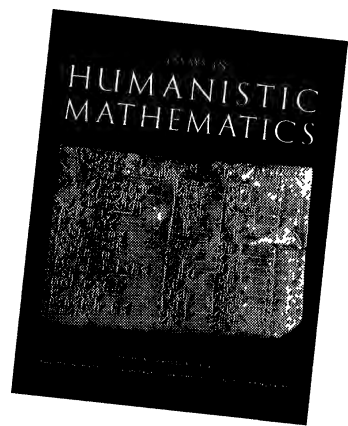


**Ensuring the future
for those who shape it.SM**

CREF certificates are distributed by TIAA-CREF Individual and Institutional Services. For more complete information, including charges and expenses, call 1 800-842-2733, ext. 8016 for a prospectus. Read the prospectus before you invest or send money.

Essays in Humanistic Mathematics

Alvin White, Editor



A dazzling array of essayists reveal humanistic mathematics in this volume, and in so doing go beyond the facts, formulas, and algorithms that most students associate with mathematics to a presentation of mathematics as an intellectual discipline with a human perspective and a significant history. Humanistic mathematics challenges dogmatic teaching styles that expect students to parrot the lecturer. It demands creativity from both the teacher and student.

Teaching mathematics humanistically seeks to place the student more centrally in the position of inquirer than is generally the case, while at the same time acknowledging the emotional climate of the activity of learning mathematics. This type of teaching encourages students to learn from each other and to better understand mathematics as socially constructed knowledge, rather than as an arbitrary discipline.

Teaching humanistic mathematics brings the focus less upon the nature of the teaching and learning environment and more upon the need to reconstruct the curriculum and the discipline of mathematics itself. This reconstruction relates mathematical discoveries to personal courage, discovery to verification, mathematics to science, truth to utility, and

mathematics to the culture in which it is embedded.

The humanistic mathematics movement, which began as the personal vision of a few, has now become a major part of mathematical culture. What was viewed with skepticism is now accepted and expected. Humanistic mathematics is not a new discovery. It is a recent rediscovery of ideas that go back to Plato. It has provided a vocabulary for previously unarticulated concepts and approaches.

The essays in this volume illustrate and help to define humanistic mathematics. The variety and scope indicate the richness and fruitfulness of the concept. Although each essay is independent, a sense of unity emerges. A glimpse at the table of contents will give you an idea of the excitement and range of the ideas presented.

212 pp., Paperbound, 1993

ISBN 0-88385-089-3

List: \$24.00

Catalog Number NTE-32

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty.	Catalog Number	Price
------	----------------	-------

_____	_____	_____
-------	-------	-------

_____	_____	_____
-------	-------	-------

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

Mathematical Cranks

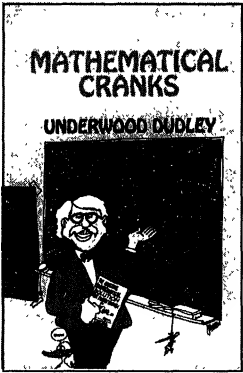
Underwood Dudley

A delightful collection...It is hard to put down and provides topics for an unending series of interesting discussions. The organization and breadth of the book are impressive, supported by a helpful index and a list of resources that encourage further explorations. A classic. —CHOICE

A jewel...The most interesting book that I have read this year.
—Journal of Recreational Mathematics

It's a gem...Dudley hasn't tossed out crank submissions over the years; he's saved them and collected samples from other mathematicians. And what wonderful samples they are.
—Sunday Telegraph, Nashua, New Hampshire

Mathematical Cranks is about people who think that they have done something impossible, like trisecting the angle, squaring the circle, duplicating the cube, or proving Euclid's parallel postulate; people who think they have done something that they have not, like proving Fermat's Last Theorem, verifying Goldbach's Conjecture, or finding a simple proof of the Four Color Theorem; people who have eccentric views, from mild (thinking we should count by 12s instead of 10s) to crazy (thinking that second-order differential equations will solve all problems of economics, politics, and philosophy); people who pray in matrices; people who find the American Revolution ruled by the number 57; people who have in common something to do with mathematics and something odd, peculiar, or bizarre.



Cranks and their ideas come in great variety. The book is a collection of examples, designed to give readers an idea of what cranks do and how they do it. Contemplating the odd, peculiar, or bizarre can be entertaining or enlightening. There can be no solution to the problem of mathematical cranks—obsessive people we will always have with us, and some will become obsessed with mathematics—but perhaps viewing the futility of their efforts will turn some prospective cranks toward more fruitful endeavors.

This is a truly unique book, written with wit and style.

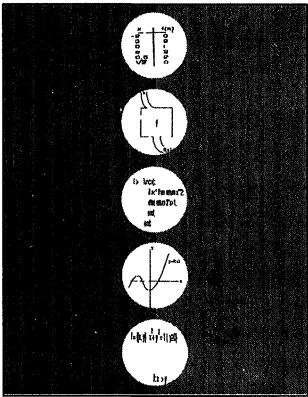
300 pp., 1992, Paperbound
ISBN 0-88385-507-0
List: \$27.50 MAA Member: \$19.50
Catalog Number CRANKS

ORDER FROM:
The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

Membership Code -----	Qty.	Catalog Number	Price
Name _____	_____		
Address _____	_____		
City _____	Total \$ _____		
State _____ Zip Code _____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
	Credit Card No. _____		
	Signature _____ Exp. Date _____		

The Concept of Function Aspects of Epistemology and Pedagogy

Guershon Harel and Ed Dubinsky, Editors



Highly readable contribution to the literature on learning the concept of function. The overall quality of the papers is quite high.
—Mathematics Teacher

The contributors of this volume probe the idea of what it means to learn the concept of function and how instruction, based on research, could assist teachers in finding ways of helping their students understand this all-important mathematical concept.

The concept of function is one that will appear again and again in a student's mathematics training. Arithmetic in the early grades, algebra in junior high school, and transformational geometry in high school are all largely based on the idea of function. Moreover, people involved in calculus reform know that understanding the idea of function is an indispensable part of the background students need to understand calculus. As mathematical education is being renewed and reformed throughout the world, this movement requires that we learn more about the concept of function both from epistemological and pedagogical points of view.

There are several major themes that emerge in the pages of this volume. They are theoretical perspectives of development of the function concept, theory-based teaching experiments, conceptions held by students and teachers, and the use of pedagogical software. The volume begins with a summary and overview of the subject and is followed by a brief glossary of terms.

The development of the papers presented in the volume began with a conference held in West Lafayette, Indiana in October 1990 with the support of Purdue University and the Exxon Foundation. This volume is, however, much more than just a conference proceedings. It is a truly cooperative writing effort by a group of dedicated researchers and educators.

350 pp., 1992, Paperbound
ISBN 0-88385-081-8
List: \$22.00
Catalog Number NTE-25

ORDER FROM:
The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

-----		-----	
Membership Code	Qty.	Catalog Number	Price
-----	-----		
Name	-----		
Address	-----		
City	-----		
State	Zip Code	Total \$	
-----	-----	-----	
Payment <input type="checkbox"/> Check		<input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD	
Credit Card No.		-----	
Signature		Exp. Date	
-----		-----	

EXCURSIONS IN CALCULUS: an Interplay of the Continuous and the Discrete

Robert M. Young

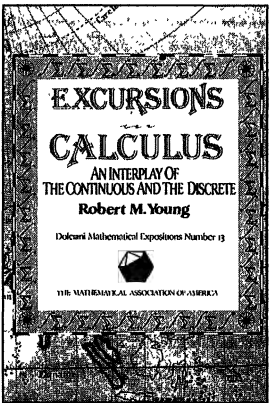
An excellent source of projects for well motivated students. This list of 463 references is a valuable aid for those who wish to dig deeper. —CHOICE

The presentation is clear and the topics very interesting... fully accessible to students for whom the book is intended. The book will be influential in awakening students' awareness for good classical mathematics. —Paulo Ribenboim

Printed with eight full-color plates.

The purpose of this book is to explore, within the context of elementary calculus, the rich and elegant interplay that exists between the two main currents of mathematics, the continuous and the discrete. Such fundamental notions in discrete mathematics as induction, recursion, combinatorics, number theory, discrete probability, and the algorithmic point of view as a unifying principle are continually explored as they interact with traditional calculus. The interaction enriches both.

The book is addressed primarily to well-trained calculus students and their teachers, but it can serve as a supplement in a traditional calculus course for anyone who wants to see more.



CONTENTS:

- Infinite Ascent, Infinite Descent: The Principle of Mathematical Induction
- Patterns, Polynomials, and Primes: Three Applications of the Binomial Theorem
- Fibonacci Numbers: Function and Form
- On the Average
- Approximation: from Pi to the Prime Number Theorem
- Infinite Sums: A Potpourri

The problems, taken for the most part from probability, analysis and number theory, are an integral part of the text. Many point the reader toward further excursions. There are over 400 problems presented in this book.

408 pp., 1992, Paperbound
ISBN 0-88385-317-5
List: \$42.00 MAA Member \$34.00
Catalog Number DOL-13

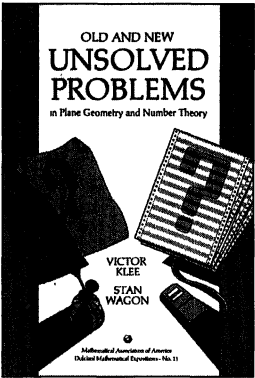
ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

Membership Code	Qty.	Catalog Number	Price
-----	_____	_____	_____
Name _____	_____	_____	_____
Address _____	_____	_____	_____
City _____	_____	_____	Total \$ _____
State _____ Zip Code _____	_____	_____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
_____	_____	_____	Credit Card No. _____
_____	_____	_____	Signature _____ Exp. Date _____

OLD AND NEW UNSOLVED
PROBLEMS IN PLANE
GEOMETRY AND
NUMBER THEORY

Victor Klee and Stan Wagon



Many facts and problems to fascinate both on familiar and unfamiliar topics. It is compulsive reading and will fill your mind with problems that will come back to haunt you again and again during idle moments.

—Mathematical Intelligencer

The book will serve well as a point of entry for students who want to know more about celebrated questions, or simply take in the vistas.

—CHOICE

This is a book that not only belongs in every university, college and high school library, it very definitely belongs in every public library.

—Mathematical Reviews

Part of the broad appeal of mathematics is that there are simply stated questions that have not yet been answered. These questions are plentiful in the areas of plane geometry and number theory, and the purpose of this book is to discuss some unsolved problems in these fields. Because the central concepts of geometry and number theory are understood by everyone, many of the questions can be understood by readers with extremely little mathematical background.

The authors place each problem in its historical and mathematical context. Each problem section is presented in two parts: The first gives an

elementary overview discussing the history and both solved and unsolved variants of the problem. Part Two contains more details, including a few proofs of related results, a wider and deeper survey of what is known about the problem and its relatives, and a large collection of references. Both parts contain exercises, and solutions to the exercises are included.

The book is aimed at both teachers and students of undergraduate mathematics, and at beginning graduate students. It could be used as a text in a course about unsolved problems, and also in courses in geometry or number theory. High school teachers interested in learning about developments in modern mathematics will find much of interest here.

352 pp., Paperbound, 1991
ISBN 0-88585-315-9
List: \$28.00 MAA Member: \$21.00
Catalog Number DOL-11

ORDER FROM:

Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC. 20036

1-800-331-1622 (FAX) (202) 265-2384

Membership Code _____	Quantity	Title	Price
Name _____	_____		
Address _____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
City _____ State _____ Zip _____	Credit Card No. _____	TOTAL \$ _____	
	Signature _____	Exp. Date _____	

Knot Theory

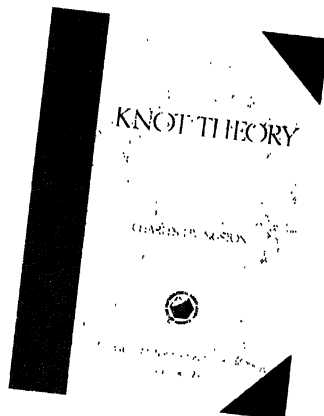
Charles Livingston

I learned more about knots after an hour with the book than I thought I could, and I am glad that it is here on my desk so that I may spend more time with it and, I hope, learn more. —Paul Halmos

Knot Theory, a lively exposition of the mathematics of knotting, will appeal to a diverse audience from the undergraduate seeking experience outside the traditional range of studies to mathematicians wanting a leisurely introduction to the subject. Graduate students beginning a program of advanced study will find a worthwhile overview, and the reader will need no training beyond linear algebra to understand the mathematics presented.

Over the last century, knot theory has progressed from a study based largely on intuition and conjecture into one of the most active areas of mathematical investigation. **Knot Theory** illustrates the foundations of knotting as well as the remarkable breadth of techniques it employs—combinatorial, algebraic, and geometric.

The interplay between topology and algebra, known as algebraic topology, arises early in the book, when tools from linear algebra and from basic group theory are introduced to study the properties of knots, including the unknotting number, the braid index, and the bridge number. Livingston guides you through a general survey of the topic showing how to use the techniques of linear algebra to address some sophisticated problems, including one of mathematics' most beautiful topics, symmetry. The book closes with a discussion of high-dimensional knot theory and a presentation of some



of the recent advances in the subject—the Conway, Jones and Kauffman polynomials. A supplementary section presents the fundamental group, which is a centerpiece of algebraic topology.

An extensive collection of exercises is included. Some problems focus on details of the subject matter; others introduce new examples and topics illustrating both the wide range of knot theory and the present borders of our understanding of knotting. All are designed to offer the reader the experience and pleasure of working in this fascinating area.

264 pp., Hardbound, 1993

ISBN 0-88385-027-3

List: \$31.50 MAA Member: \$25.00

Catalog Number CAM-24

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036
1-(800) 331-1622 (202)-387-5200

Membership Code

Name _____

Address _____

City _____

State ____ Zip Code _____

Qty. Catalog Number Price

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

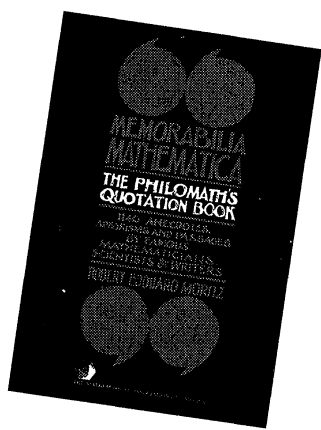
Signature _____

Exp. Date _____

Memorabilia Mathematica

The Philomath's Quotation Book

Robert Edouard Moritz



When Robert Edouard Moritz compiled his book of quotations, **Memorabilia Mathematica**, which appeared in 1914, he stated that his primary objective was to seek out the exact statement of and exact references for famous passages about mathematics. He searched the writing not only of mathematicians, but poets, philosophers, historians, statesmen, and scientists as well. His sources ranged from the works of Plato to the writings of Hilbert and Whitehead. His second objective was to produce a volume that would be a source of pleasure, encouragement, and inspiration to both mathematicians and non-mathematicians alike.

This work was a ten-year labor of love, and it is a tribute to his discerning eye that this selection of passages should remain one of the most stimulating works about mathematics ever published. It was the first collection of its kind in English and it conveys a sense of the full range of mathematics, its enormous accomplishments, and the living personalities of great mathematicians.

The more than eleven-hundred fully annotated selections in this book, gathered from the works of three hundred authors, cover a vast range of subjects pertaining to mathematics. Grouped in twenty-one chapters, they deal with such topics as the definitions and objects of mathematics; the teaching of mathematics; mathematics as a language or as a fine art; the relationship of mathematics to philosophy, to logic, or to science; the

nature of mathematics, and the value of mathematics. Other sections contain passages referring to specific subjects in the field such as arithmetic, algebra, geometry, calculus, and modern mathematics. Of special interest is the extensive amount of material on great mathematicians which provides irreplaceable glimpses into the lives and personalities of mathematical giants.

To mathematicians the book will be a great source of pleasure, inspiration, and encouragement. To teachers of mathematics and writers about mathematics, it will remain of inestimable value as a source of quotations and ideas. To the layperson, it will be a revelation. It should dispel forever the narrow notion that mathematics is a cut-and-dried affair, isolated from other compartments of life and thought.

440 pp., Paperbound, 1993

ISBN 0-88385-321-3

List: \$24.00 MAA Member: \$19.00

Catalog Number: MEMO

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Name _____

Address _____

City _____

State ____ Zip Code _____

Qty.	Catalog Number	Price
------	----------------	-------

_____	_____	_____
-------	-------	-------

_____	_____	_____
-------	-------	-------

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

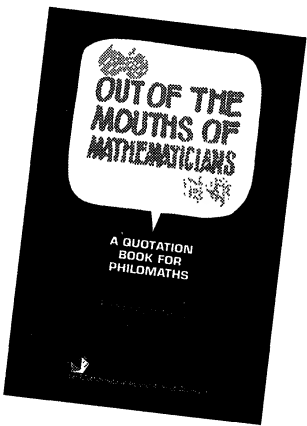
Out of the Mouths of Mathematicians

A Quotation Book for Philomaths

Rosemary Schmalz

Published as a companion volume to Robert Edouard Moritz's *Memorabilia Mathematica*, Rosemary Schmalz's *Out of the Mouths of Mathematicians* picks up where Moritz left off. Her work will give you a sense of the "story" of twentieth century mathematics. You will encounter the mathematicians, their collaborations and disputes, the movement from abstraction to application, the emergence of new areas of research, the impact of computers on mathematics, the challenges in mathematics education, and more.

Out of the Mouths of Mathematicians: A Quotation Book for Philomaths is a compilation of 727 quotations from 292 contributors, almost all of whom are twentieth century mathematicians. Taking the advice of Abel to learn mathematics by reading the masters, the author offers the reader a unique perspective on this century's mathematics through the words of the mathematicians who are its creators. Stories about these mathematicians, their exhortations to their students, their descriptions of their efforts, successes, and failures, all make this century's mathematics come alive. The book also offers readers the opportunity to broaden their ideas about what mathematics is by offering many definitions of mathematics, making comparisons of mathematics to computing and to the fine arts, and showing similarities between many aspects of mathematics and religion. The complete reference for each quotation allows the reader to continue exploration into a favorite area. A large topic index makes the book quite user-friendly. Some of the subject categories include:



The Development of Mathematics, Exhortations to Aspiring Mathematicians, Pure and Applied Mathematics, About Mathematicians (by name), Anecdotes and Miscellaneous Humor, Particular Disciplines in Mathematics, Moments of Mathematical Insight, Mathematics and the Arts,... and much more.

This book will give pleasure to any philomath. It can be used to facilitate a literature search or to give quick access to an appropriate quote for writers and speakers. It will be particularly useful to teachers of mathematics at all levels, to encourage, motivate, and amuse their students. Along with R. E. Moritz's earlier book of this type, *Memorabilia Mathematica: The Philomath's Quotation Book*, it offers the story of mathematics from its primary source, the mathematicians themselves.

304 pp., Paperbound, 1993
ISBN 0-88385-509-7
List: \$29.00 MAA Member: \$23.00

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Name _____
Address _____
City _____
State _____ Zip Code _____

Catalog Number OMMA

Qty.	Catalog Number	Price
Total \$		
Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
Credit Card No. _____		
Signature _____		
Exp. Date _____		

The Search for E.T. Bell

also known as John Taine

Constance Reid

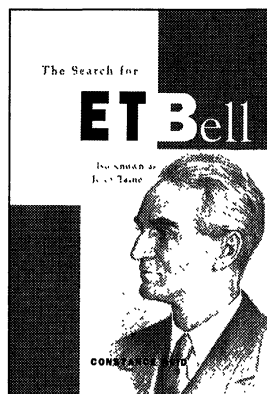
No one today writes about mathematics and mathematicians with more grace, knowledge, skill, and clarity, and no one is going to produce a more delightful, informative, accurate account of Eric Temple Bell and his work, and that of his alter-ego, the prolific pioneer of science fiction, John Taine. This is a fine book. —Martin Gardner

Eric Temple Bell has been one of my heroes for 60 years...I congratulate Constance Reid on a remarkable achievement. I hope it is greeted with the success it deserves, and revives interest in an extraordinary and multi-talented man.

—A. C. Clarke

Eric Temple Bell (1883–1960) was a distinguished mathematician and a best selling popularizer of mathematics. His *Men of Mathematics*, still in print after almost sixty years, inspired scores of young readers to become mathematicians. Under the name “John Taine,” he also published science fiction novels (among them *The Time Stream*, *Before the Dawn*, and *The Crystal Horde*) that served to broaden the subject matter of that genre during its early years.

In *The Search for E.T. Bell*, Constance Reid has given us a compelling account of this complicated, difficult man who never divulged to anyone, not even to his wife and son, the story of his early life and family background. Her book is thus more of a mystery than a traditional biography. It begins with the discovery of an unexpected inscription in an English churchyard and a series of cryptic notations in a boy's schoolbook. Then comes an inadvertent revelation, by Bell himself, in a respected mathematical journal... You will have to read the book to learn the rest.



Originally agreeing to write only a profile of Bell, Mrs. Reid soon found herself involved in a full-length biography. The discoveries she made in the course of her five years of research will necessitate a fresh evaluation of his extensive mathematical work and his science fiction novels as well as the revision of almost every statement currently in print about his family background and early life. Mrs. Reid is already well known as the author of acclaimed biographies of David Hilbert, Richard Courant, and Jerzy Neyman.

Includes a collection of over 75 photographs.

384 pp., Hardbound, 1993

ISBN 0-88385-508-9

List: \$35.00 MAA Member: \$28.00

Catalog Number BELL

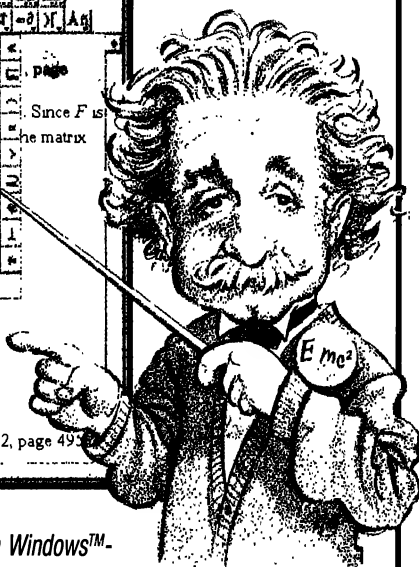
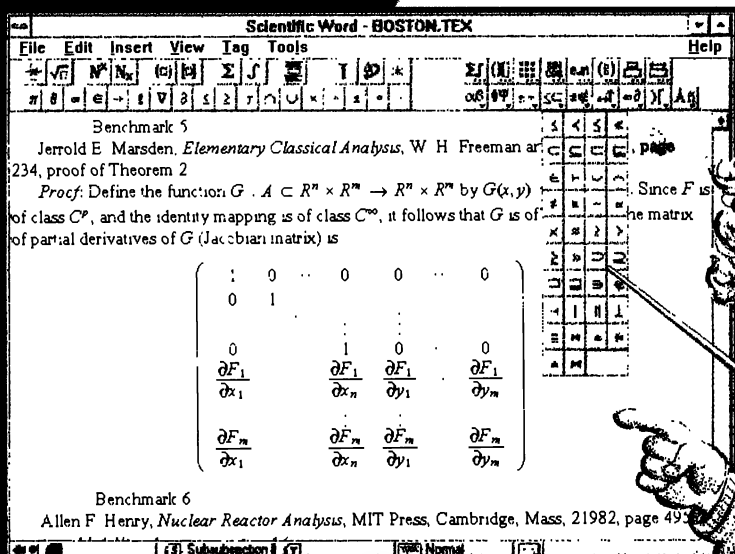
ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
(202) 387-5200 1-(800)331-1622

Membership Code	Qty.	Catalog Number	Price

Name _____			Total \$ _____
Address _____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
City _____	Credit Card No. _____		
State ____ Zip Code _____	Signature _____		
	Exp. Date _____		

SCIENTIFIC[®] Word



Discover the Genius of Scientific Word...

Scientific Word, a Windows[™]-based front-end to L^AT_EX is easy to learn and easy to use. It is a full document processor, not just an equation editor. You enter text and mathematics on a continuous screen without the distraction of popping in and out of equation boxes. With **Scientific Word** you use familiar mathematical notation to enter your mathematics – you need no special codes. **Scientific Word** adheres to internationally accepted mathematical formatting standards, so you are free to deal with the content of your document rather than its appearance. Your document is saved as an ASCII L^AT_EX file and printed output is produced via T_EX, the mathematical typesetting standard.

Call toll free **1-800-874-2383** to order your copy today. Ask about our 30-day money-back guarantee and our educational discount.

800-874-2383

CALL TODAY FOR MORE INFORMATION!



SOFTWARE RESEARCH

1190 FOSTER ROAD
LAS CRUCES, NM 88001

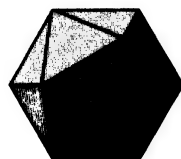
TEL: (505) 522-4600
FAX: (505) 522-0116

sales@tcisoft.com

SCIENTIFIC WORD IS A REGISTERED TRADEMARK OF TCI SOFTWARE RESEARCH.
T_EX IS A TRADEMARK OF THE AMERICAN MATHEMATICAL SOCIETY.
WINDOWS IS A TRADEMARK OF MICROSOFT.

The American Mathematical Monthly

Volume 101 Number 4 / APRIL 1994
(ISSN 0002-9890)



Contents

ARTICLES

Pizza Slicing, Φ 's and the Riemann Hypothesis / EDWARD A. BENDER,
OREN PATASHNIK, and HOWARD RUMSEY, JR. 307

Rational Periodic Points of the Quadratic Function $Q_c(x) = x^2 + c$ /
RALPH WALDE and PAULA RUSSO 318

Fréchet vs. Carathéodory / ERNESTO ACOSTA G.
and CESAR DELGADO G. 332

Odd Magic Powers / A. C. THOMPSON 339

Mathematicians, Including Undergraduates, Look at Soap Bubbles /
FRANK MORGAN 343

FEATURES

COMMENTS 306

NOTES

A Proof of Dilworth's Chain Decomposition Theorem /
FRED GALVIN 352

On Intervals, Transitivity = Chaos / MICHEL VELLEKOOP
and RAOUL BERGLUND 353

Proof of a Mixed Arithmetic-Mean, Geometric-Mean Inequality /
KIRAN KEDLAYA 355

UNSOLVED PROBLEMS

ApSimon's Mints Problem 358

THE AUTHORS 360

PROBLEMS AND SOLUTIONS 362

REVIEWS

*Mathematics of the 19th Century: Mathematical Logic, Algebra,
Number Theory, Probability Theory*, edited by A. N. Kolmogorov
and A. P. Yushkevich / KAREN HUNGER PARSHALL 369

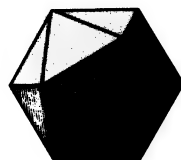
Calculus Gems: Brief Lives and Memorable Mathematics.
By George F. Simmons / DAVID J. PENGELLEY 374

TELEGRAPHIC REVIEWS 381

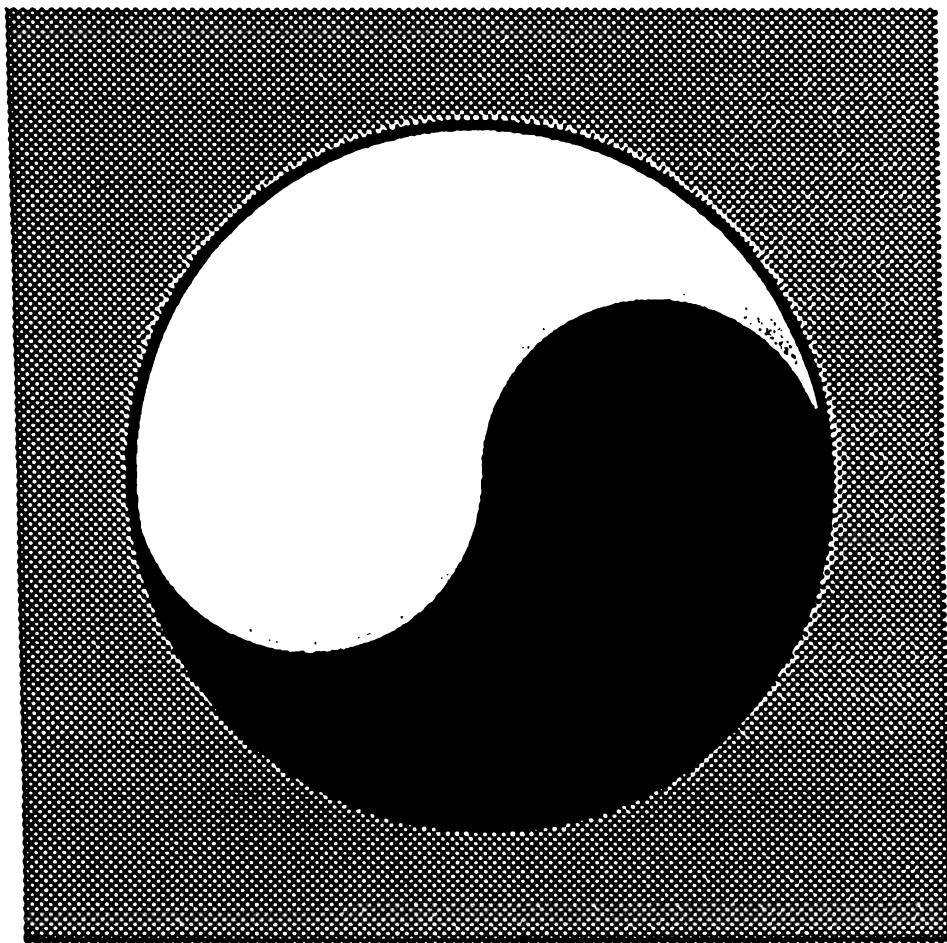
THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, D.C. 20036



The American Mathematical Monthly



Volume 101, Number 5 / MAY 1994



AN OFFICIAL PUBLICATION OF THE MATHEMATICAL ASSOCIATION OF AMERICA

NOTICE TO AUTHORS

The *Monthly* publishes articles, notes, and other features about mathematics and the profession. The readership of the *Monthly* is intended to include everybody who is mathematically inclined, including of course professional mathematicians and students of mathematics at all collegiate levels. While no single article or feature is likely to appeal to everyone, material should interest and be accessible to a large number of readers. This is the most important criterion for acceptance.

Articles may be expositions of old results or presentations of new ones. They may concern all of mathematics or one small area, a broad development or a single application, historical reminiscences or one important event. While some articles may contain the author's new research, the novelty of material and generality of the results is far less important than the clarity of exposition and general interest. Discussing one illuminating case of a well known result is far better than providing all the details of an obscure but new proposition. Articles in the *Monthly* are supposed to inform and to entertain; they are meant to be read rather than archived.

Notes are short and possibly informal articles. A note may concern a clever new proof of an old theorem, a novel way to present tired material, or a lively discussion of a philosophical (but still mathematical) issue. Also, any topic is suitable, so long as it is related to mathematics. Because a note is short, the first few sentences are the most important part: They should explain the purpose and invite the reader in. Photographs or diagrams often will attract the reader's attention.

All articles and notes should be sent to the editor:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Please send 3 copies, typewritten on only one side of the paper. Illustrations should be carefully drawn on separate sheets of paper in black ink; the original should be without lettering and two copies should have appropriate captions and lettering indicated.

Proposed problems or solutions should be sent to:

RICHARD BUMBY,
P.O. Box 1971
New Brunswick, NJ 08906-0971.

Please send 2 copies of all material, typewritten if possible.

Letters to the Editor, both for publication and for private reading, should be sent to the Editor at the address given above. Comments, including criticisms, are welcome, as are all suggestions for making the *Monthly* a lively, entertaining, and informative journal.

EDITOR:

JOHN H. EWING

ASSOCIATE EDITORS:

PETER BORWEIN	FRED KOCHMAN
RICHARD BUMBY	CATHERINE MCGEOCH
DENNIS DETURCK	RICHARD NOWAKOWSKI
UNDERWOOD DUDLEY	ARNOLD OSTELEE
JOHN DUNCAN	LEE RUBEL
JOAN FERRINI-MUNDY	ABE SHENITZER
JOSEPH GALLIAN	LYNN STEEN
STEVEN GALOVICH	STAN WAGON
RICHARD GUY	DOUGLAS WEST
DARRELL HAILE	HERBERT WILF
PAUL HALMOS	SANDY ZABELL
JOAN HUTCHINSON	PAUL ZORN

EDITORIAL ASSISTANT:

MISTY CUMMINGS

STAFF ARTIST:

MIKE CAGLE

Reprint permission:

MARCIA P. SWARD, Executive Director

Advertising Correspondence:

Ms. ELAINE PEDREIRA, Advertising Manager

Subscription correspondence, change of address, and other inquiries:

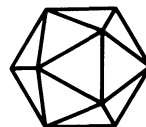
Membership / Subscriptions Department

All at the address:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036.

Microfilm Editions: University Microfilms International,
Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT. Copyrighted by the Mathematical Association of America (Incorporated), 1994, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source. Second class postage paid at Washington, DC, and additional mailing offices. **Postmaster:** Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.



Contents

ARTICLES

**On the Geometry of Piecewise Circular Curves / THOMAS BANCHOFF
and PETER GIBLIN 403**

The Two Envelope Paradox / ELLIOT LINZER 417

Fourier Series of Polygons / ALAIN ROBERT 420

The Paradox of Nontransitive Dice / RICHARD P. SAVAGE, JR. 429

**Squares Expressible as Sum of Consecutive Squares /
LAURENT BEECKMANS 437**

Square Roots mod p / STEPHEN M. TURNER 443

FEATURES

COMMENTS 402

NOTES

**Kummer's Test Gives Characterizations for Convergence or Divergence
of all Positive Series / JINGCHENG TONG 450**

Isometries of ℓ_p -norm / CHI-KWONG LI and WASIN SO 452

**A Trace Inequality for Unitary Matrices / BOYING WANG
and FUZHEN ZHANG 453**

**An Elementary Proof of the Square Summability of the Discrete Hilbert
Transform / LOUKAS GRAFAKOS 456**

THE COMPUTER SCIENCE SAMPLER

**Does Anybody Really Know What Time It Is? /
CATHERINE C. MCGEOCH 459**

THE EVOLUTION OF...

**How Hyperbolic Geometry Became Respectable /
ABE SHENITZER 464**

THE AUTHORS 471

PROBLEMS AND SOLUTIONS 473

REVIEWS

***The Lure of the Integers.* By Joe Roberts / PAUL T. BATEMAN
and HAROLD G. DIAMOND 480**

***Excursions in Calculus: An Interplay of the Continuous and the Discrete.*
By Robert M. Young / ANITA E. SOLOW 482**

TELEGRAPHIC REVIEWS 485

COMMENTS

It is not the employer who pays wages — he only handles the money. It is the product that pays wages.

--Henry Ford

Rebates.

It's the free enterprise solution both to grade inflation and to the lack of accountability. Here's how it works. Students pay an exorbitant fee for each credit hour or course. How high? Triple whatever students pay now at your institution. (So far, this sounds normal — at least for a parent with children in college). Now comes the difference. Depending on their grade in the course, students receive *rebates*. An A returns a whopping 80%; a B gives 60%; a C 40%; a D 20%; and an F gives nothing at all. Rebate checks (along with the grade report) are mailed to expectant parents promptly at the end of each term.

What happens to the money generated by higher tuition? A portion is returned to each unit involved in the instruction — the college, the department, the instructor — perhaps even the grading assistant.

That's it, the system; set it up and let it run.

This is a system with all the right economic forces . . . and with real *accountability*. Mom and Dad can pay for their winter cruise when they receive a check at the end of good semester. (It's more tangible than a small slip of paper with 3.8 typed on it.) Students who pay their own way will be even happier. Money is motivation.

What about grade inflation? Too many A's and B's and the college and department soon go broke. Too few A's and B's means no one signs up for the courses in that expensive department. For instructors, the incentive is even clearer: Grade too easily and you're poor, grade too hard and no one takes your courses (and you're still poor). Clearly there exists a maximum profit in between.

There is flexibility in this system. When the departmental xerox machine breaks down in December, an immediate call goes out to make finals tougher. They can raise the funds with one demanding Calculus test. On a smaller scale, faculty can generate the down payment on a new house, or the cost of an unexpected auto repair, by teaching one large course with a tough curve. Don't do it too often, however, or you won't have the chance again.

The system has an additional advantage for people who like to use simple numbers to measure young faculty at tenure time. "Mary has only a 3.6 in teaching evaluations," they point out, "but she's generated \$47,563 for the department in two years."

Above all, this is a system that recognizes the proper relation of students as *customers* and teachers as *vendors*. The product we sell pays our wages.

Does all this make you uneasy?
I hope so.

John Ewing

On the Geometry of Piecewise Circular Curves

Thomas Banchoff and Peter Giblin

In this article we would like to promote a class of plane curves that have a number of special and attractive properties, the piecewise circular curves, or PC curves. (We feel constrained to point out that the term has nothing to do with Personal Computers, Privy Councils, or Political Correctness.) They are nearly as easy to define as polygons: a *PC curve* is given by a finite sequence of circular arcs or line segments, with the endpoint of one arc coinciding with the beginning point of the next. These curves are more versatile than polygons in that they can have a well-defined tangent line at every point: a PC curve is said to be *smooth* if the directed tangent line at the end of one arc coincides with the directed tangent line at the beginning of the next. (In particular, in a smooth PC curve, no arc degenerates to a single point.)

In the literature of descriptive geometry and more recently in computer graphics, PC curves have been used to approximate smooth curves so that the approximation is not only pointwise close, as in the case of an inscribed polygon, but also has the property that the tangent lines at the points of the smooth curve are approximated by the tangent lines of the PC curve. Given a pair of nearby points on a smooth curve together with their tangent directions, there will not in general be a single circular arc through the points with those directions at its endpoints, but there will be a family of *biarcs* meeting these boundary conditions, PC curves composed of two tangent circular arcs. (See [M-N] for a discussion of this construction.)

EXAMPLES OF PC CURVES. PC curves arise naturally as the solutions of a number of variational problems related to isoperimetric problems. A classical problem is to find the curve of shortest length enclosing a fixed area, and the solution is a circle. If the curve is required to surround a fixed pair of points, then the curve of shortest length enclosing a given area will be either a circle or a *lens* formed by two arcs of circles of the same radius meeting at the two points. More generally Besicovitch has shown that a curve of fixed length surrounding a given convex polygon and enclosing the maximum area must be a PC curve with all radii of arcs equal [Be]. One such curve is the Reuleaux “triangle”, a three-arc PC curve enclosing an equilateral triangle, with each radius equal to the length of a side of the triangle. Such three-arc PC curves, and many far more elaborate examples can be found in the tracery of gothic windows [A].

If we require that a curve of fixed length L surround a given pair of discs of the same radius, then, for a certain range of values of L , the curve that encloses the greatest area is a smooth convex PC curve consisting of two arcs on the boundary circles of the discs and two arcs of equal radius tangent to both discs. Such four-arc convex PC curves have long been used in engineering drawing for approximating ellipses, and we call such a curve a *PC ellipse* [FIGURE 1]. One special PC ellipse is

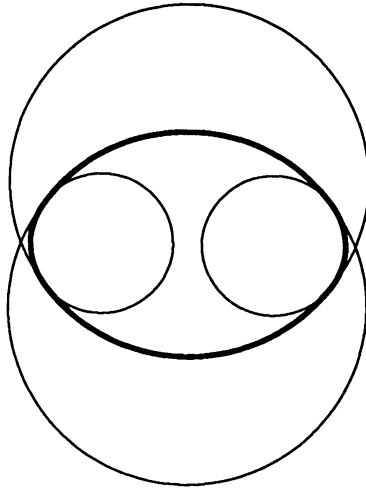


Figure 1

the boundary of the smallest convex set containing the two discs, called the *convex envelope* of the two discs, consisting of two semicircles and two line segments. (We thank Salvador Segura for pointing out the importance of PC curves in such isoperimetric problems.)

The collection of PC curves is invariant not only under Euclidean motions and scaling, but also under inversion with respect to a circle.

PARALLEL CURVES OF PC CURVES. The Reuleaux triangle is a non-smooth PC curve of *constant width*, so that every strip containing the curve and bounded by a pair of parallel lines through points of the curve has the same width. [FIGURE 2a]. We can obtain a smooth PC curve of constant width by taking an *outer parallel curve* of the Reuleaux triangle, i.e. the boundary of the parallel region, the locus of all points within a fixed distance of the points of Reuleaux triangle [FIGURE 2b].

This construction points out one of the main properties of PC curves: since the parallel curves of circular arcs are circular arcs, the parallel curves of a PC curve

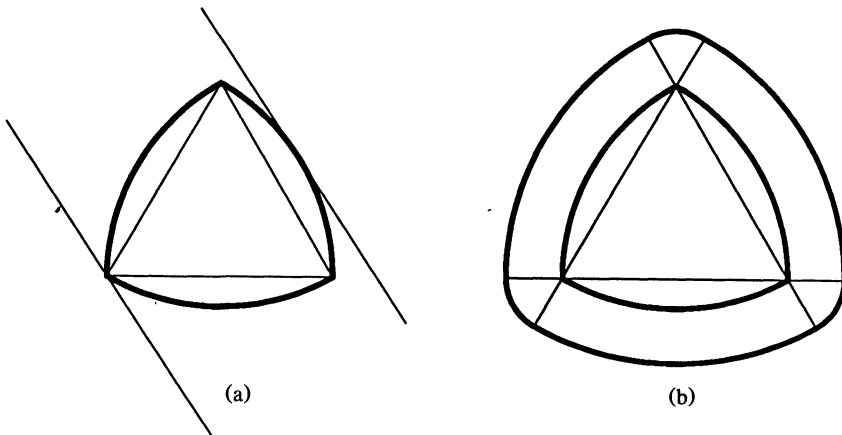


Figure 2

are PC curves. As an example, consider a convex PC ellipse. If we increase all of the radii by the same amount, keeping the same centers for the arcs, we obtain an outer parallel curve which is also a PC ellipse. This situation is in contrast with the case of an actual ellipse, for which the exterior parallel curves are not conic sections but rather algebraic curves of fourth degree.

Consideration of parallel curves is especially important in computer graphics, in particular in robotics, where it is necessary to find the centers of all discs of a fixed radius touching a given curve. In this subject, parallel curves are often called *offset curves*, obtained by moving away from the curve a given distance. For a curve defined by an algebraic equation, the offset curves are also algebraic, but the degrees of the offset curve is in general much higher [R-R].

If instead of increasing all radii of arcs of a PC ellipse, we decrease all radii by the same amount, keeping the same centers, we obtain the family of *inner parallel curves*. As in the case of the ordinary ellipse, for sufficiently small radius, the inner parallel curves remain smooth, and in the PC case, they remain PC ellipses. For an ellipse, after a certain distance the inner parallel curve develops four *cusps* where the directed tangent line reverses direction. Similarly, at a distance equal to the smaller radius, the parallel PC curve degenerates into a lens, and just after this we obtain a four-arc PC curve with cusps, where two arcs come together at the same tangent line but with different directions. Beyond a certain distance, the parallel curve of an ellipse is again a convex curve (but not a conic section). For a convex PC ellipse, the inner parallel curve at distance equal to the larger radius is a lens, and after that, it is again a convex PC ellipse.

EVOLUTE POLYGONS OF PC CURVES. In the case of an ellipse, the cusps of parallel curves trace out the *evolute curve*, consisting of the locus of centers of curvature of the ellipse. For a PC ellipse, the cusps of the parallel curves trace out the edges of a polygon with vertices at the centers of the arcs, called the *evolute polygon* of the PC curve [FIGURE 3]. If the radius of the inner parallel curve equals

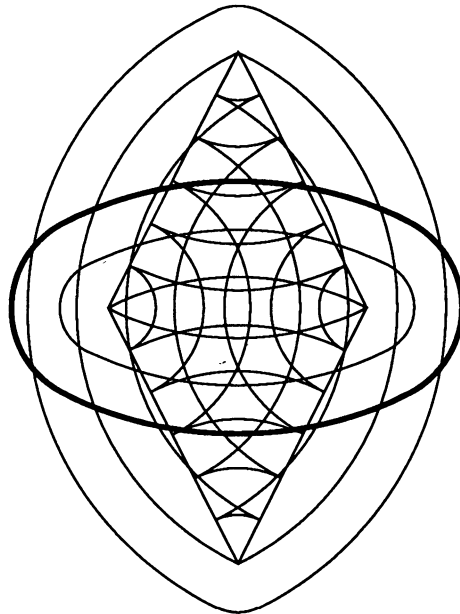


Figure 3

the radius of one of the arcs of a PC curve, then that arc degenerates to a single point. We say that a PC curve is *non-degenerate* if all arcs have non-zero length. Almost all parallel curves of a PC curve are non-degenerate.

We can obtain smooth PC curves by starting with a sequence of circles, each one tangent to its successor. The points of tangency divide each circle into two arcs, and choosing one arc from each circle gives a PC curve with a well-defined tangent line at each *node* where two successive arcs meet. If we wish the resulting PC curve to be smooth, then once we have chosen an arc from the first circle, the arcs on all subsequent circles are uniquely determined. If the last circle is tangent to the first, then this construction gives a *closed* PC curve. If each of the circles is externally tangent to the next, then the resulting PC curve which will be smooth if the number of arcs is even [FIGURE 4], while if the number of arcs is odd, then we inevitably obtain a cusp when we return to the starting point.

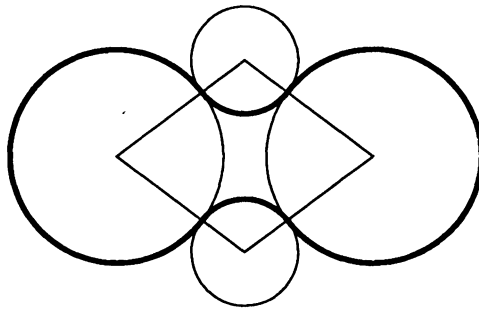


Figure 4

If two successive circles are externally tangent, then the node of the PC curve will either be a smooth *inflection point* [FIGURE 5a] if the tangent lines have the same direction or an *ordinary cusp* [FIGURE 5b] if the directions are different. In each of these cases, the two arcs lie on opposite sides of their common tangent line at the node. If two circles are internally tangent, then the two arcs at the node lie on the same side of their common tangent line, and we obtain either a *smooth locally convex point* [Figure 5c] if the tangent lines have the same direction or a *ramphoid cusp* [FIGURE 5d] if the directions are different.

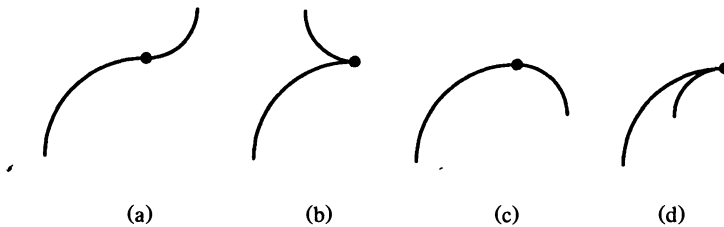


Figure 5

Once we begin to consider curves with cusps, we can obtain many of them from the same collection of successively tangent circles. Selecting one of the two possible arcs of each of the circles, we get 2^n such curves if there are n circles. If n is odd, we obtain two such curves with cusps at all nodes [FIGURE 6a–d].

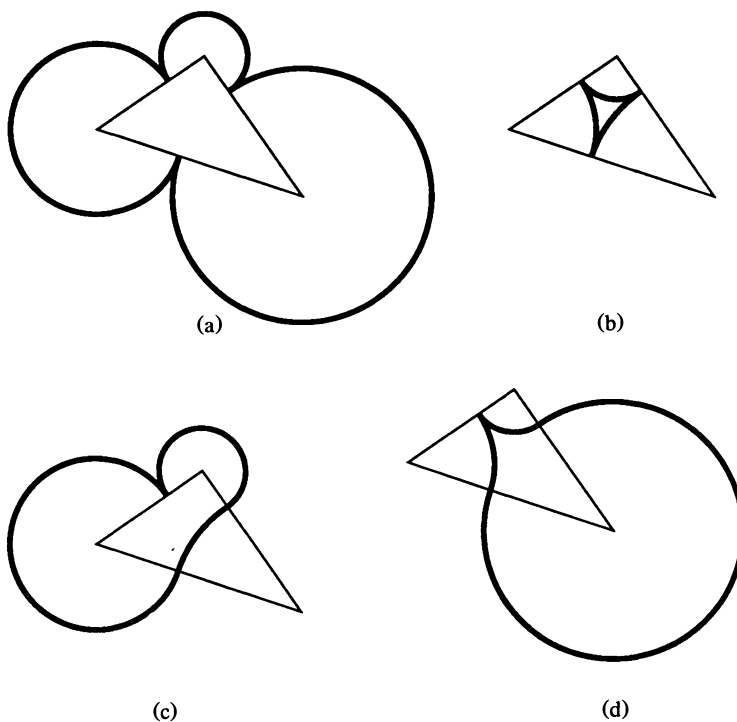


Figure 6

THREE-ARC PC CURVES. If two of the circles are inside a third, we can form eight PC curves in this way, leading to three distinct types, each of which appears in a classical guise. If the inner circles have half the radius of the outer one, then one such curve is the Yin-Yang curve, with one convex smooth node, one inflection node, and one node which is a rhamphoid cusp [FIGURE 7a]. From the same set of circles we can form the *PC cardioid* with two smooth convex points and one ordinary cusp. This curve appears in the work of the eighteenth century Jesuit geometer Roger Boscovich as an example of a non-centrally symmetric curve with

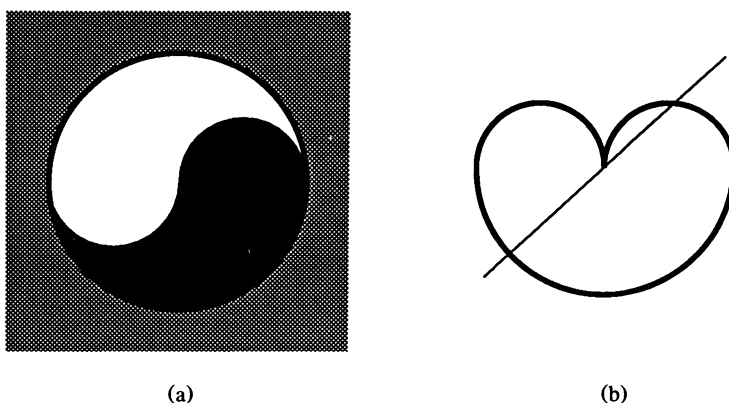


Figure 7

a *center of length*, so that every line through this center cuts the curve into two pieces of equal length [FIGURE 7b]. A third type of PC curve determined by these three circles is the *arbelos*, or shoemaker's knife, with two rhamphoid cusps and one ordinary cusp. This curve was originally studied by Archimedes and Pappus, and it has inspired numerous articles in recreational mathematics, for example [Ba], [Ga], and [H].

EVOLUTES, INVOLUTES, AND OSCULATING CIRCLES. The sequence of centers of the circular arcs of a PC curve determines the *evolute polygon* of the curve. For a PC ellipse, the evolute polygon is a rhombus, and we can find non-convex PC curves with the same rhombus as its evolute polygon [FIGURE 4 and FIGURE 8]. Any parallel curve of either of these PC curves will be a PC curve with the same evolute polygon.

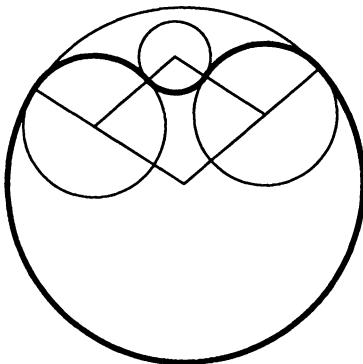


Figure 8

If we start with a sequence of circles, each one internally tangent to its successor and contained within it, then we obtain a *PC spiral*. The evolute polygon of the spiral will be a locally convex polygonal arc, and we may recover the spiral by a “string construction”. We think of a string attached at one end of the polygonal arc and pulled tightly along it. As we unwind the string, keeping it tightly along the polygon at all times, the endpoint of the string traces out a PC curve with the polygon as evolute polygon [FIGURE 9]. By using a longer string, we may construct a parallel PC curve with the same evolute polygon. Such PC spirals have been used

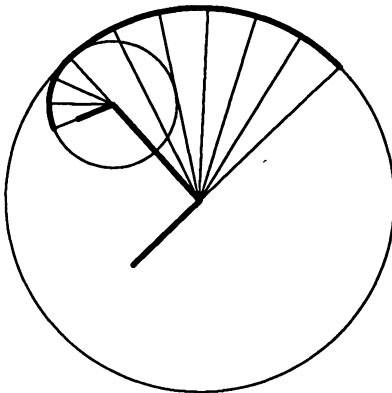


Figure 9

by several authors in computer-aided design as a means of approximating curves with increasing curvature [M-P].

For a smooth curve with continually increasing curvature, the best approximating circle at a point, called the *osculating circle* at the point, is defined by the properties that it is tangent to the curve at the point and it crosses from one side of the curve to the other near that point. The evolute curve is then the locus of centers of osculating circles at the points of the curve. At a node of a convex PC curve, the circles which are tangent to the curve at the node and which cross from one side of the curve to the other near the point have their centers on the segment joining the centers of the two arcs that meet at the node [FIGURE 10]. For this reason we may consider the evolute polygon as the locus of centers of “osculating circles” of the PC curve. At an inflection node, the circles tangent to the curve that cross from one side of the curve to the other have their centers on the line containing the centers of the arcs meeting at the node, but on the two rays that are the complement of the segment joining the two centers. In this case, the focal polygon is said to go to infinity.

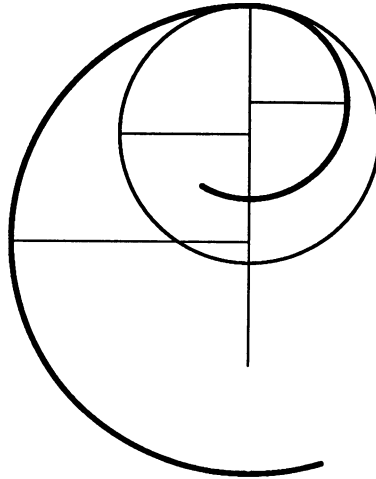


Figure 10

For a smooth spiral with continually increasing curvature, the radii of the osculating circles continually decrease, and conversely. A point where the curvature stops increasing and begins decreasing or conversely is called a *vertex*. For a PC curve, a *vertex arc* is an arc such that both adjacent arcs either are inside the circle of the arc or outside it. For a convex PC ellipse, each arc is a vertex arc.

FOUR-ARC PC CURVES. In the remainder of this article, we will discuss some results about closed four-arc PC curves, and point out an interesting connection between these and four-bar linkages in the plane. In effect, we show that all closed four-arc PC curves can be generated in a simple way from a very special class of “collapsed” quadrilaterals.

Let us establish some notational conventions. Let C_i be a circle with center c_i , $i = 1, 2, 3, 4$, each C_i being tangent to C_{i+1} . (We adopt the convention that all subscripts are to be reduced modulo 4, so for example C_4 is the same as C_0). Our

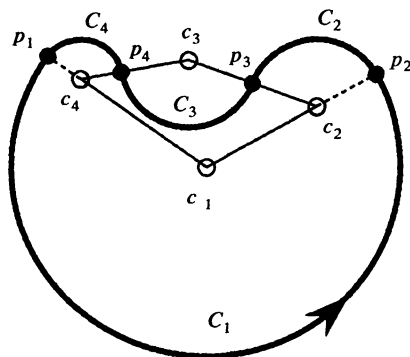


Figure 11

PC curve C will be made from successive arcs of the C_i , so that the quadrilateral $c_1c_2c_3c_4$ is the evolute polygon of C . The nodes of C will be denoted p_i , with p_i as the node where C_{i-1} meets C_i [FIGURE 11]. Finally the side-lengths of the evolute polygon will be denoted by l_i : this is the distance between c_{i-1} and c_i . In the example of FIGURE 11 it is clear, by splitting each l_i into a sum of two radii, that $l_1 + l_4 = l_2 + l_3$. Whenever we have a PC curve based on the above quadrilateral, we must have some relation of the form

$$l_4 = \pm l_1 \pm l_2 \pm l_3, \quad (1)$$

for some choice of plus or minus signs.

Suppose we start with two circles, C_1 and C_3 . What choice do we have for the centers of the remaining two circles? If, for example, the circles C_1 and C_3 are external to each other, as in FIGURE 12a, then a circle C_2 tangent to both has $|l_2 - l_3|$ equal to the sum or difference of the radii of C_1 and C_3 . Thus, by a standard property of hyperbolas, the center of C_2 (and likewise of C_4) lies on one of two hyperbolas with foci at c_1 and c_3 . For one hyperbola, C_1 and C_3 are both outside or both inside C_2 ; for the other hyperbola, one is outside and one is inside. If the radii of C_1 and C_3 are equal, then one of the hyperbolas degenerates to a straight line. If C_1 and C_3 are differently placed (for example if one is inside the other), or if one of the circles becomes a straight line (a “circle of infinite radius”), then there may be changes in the locus of possible centers for C_2 . However, as the reader may verify, this locus always consists of two *conic sections*, i.e. an ellipse, a parabola, a hyperbola, or a straight line.

A particularly interesting construction which can be carried out for any PC curve C , is to consider the full locus of centers of *bitangent circles*, i.e. circles tangent to C in at least two places. This locus, which necessarily includes the centers of the arcs making up C , is the *symmetry set* of C . By the above remarks, the symmetry set consists of arcs of conic sections; in fact, two consecutive arcs will always meet with a common tangent line. In the final section of this paper, we give an intriguing example of a four-arc PC curve which has an *isolated point* on its symmetry set; a full discussion of symmetry sets of PC curves appears in [Ba-G2].

Suppose now that we have four circles with each tangent to the next. As pointed out above, exactly two of the sixteen possible PC curves made up from arcs of these four circles are smooth.

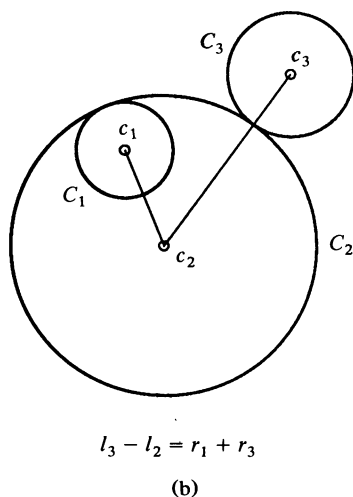
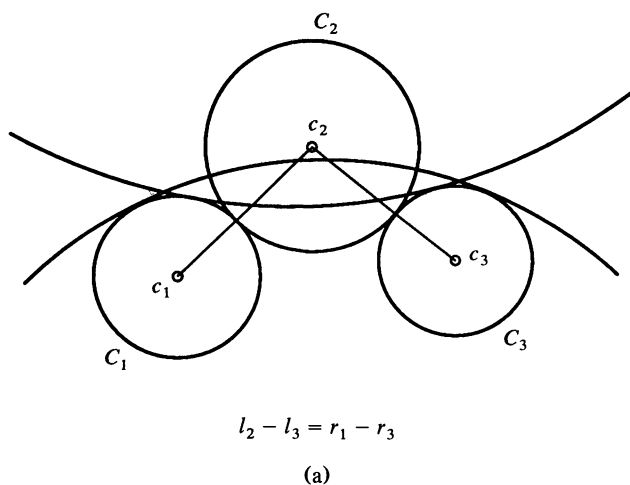


Figure 12

In the case of four externally tangent circles (as in FIGURE 4), it is clear that the arcs making up one of these smooth PC curves must alternate between clockwise and counter-clockwise orientation. With the usual convention that clockwise curves have negative curvature and counter-clockwise curves have positive curvature, this implies that each of the four arcs is a vertex arc, as defined previously. The reader may like to verify that, with configurations other than externally tangent circles, all four arcs remain vertex arcs so long as the PC curve remains smooth and does not intersect itself. This establishes a version of the Four Vertex Theorem for four-arc PC curves. More generally, one can show that any smooth closed PC curve which does not intersect itself has at least four vertex arcs. This can be proved using an argument analogous to that of Osserman, for the classical Four Vertex Theorem [O].

PC CURVES AND FOUR-BAR LINKAGES. It is instructive to regard the evolute polygon as a linkage in the plane. This amounts to thinking of the edges as rigid

rods connected at the endpoints c_i . We usually take c_1 and c_4 as fixed in position in the plane and we allow the other three rods to turn about their endpoints. As the quadrilateral changes in shape, so will any PC curve with this quadrilateral as its evolute. We may note that a collection of four tangential circles centered at the vertices c_i will roll on each other without slipping as the linkage changes shape.

The theory of four-bar linkages has been studied extensively. In that theory, it is shown that a linkage can move continuously from any position into a collapsed position, where all the centers are on a straight line (and hence all nodes lie on the same line, too), provided that some relation of the form (1) holds (where l_4 is the length of the fixed rod). In certain cases, namely

$$l_4 = l_1 + l_2 + l_3 \quad \text{and} \quad l_4 = -l_1 + l_2 - l_3,$$

there is only one possible position for the linkage and that is when it is in a collapsed position, so the result holds automatically in these cases. We state and prove the result below, using an elementary argument; for a more general setting of this result, see for example [G-N].

For us, the main significance of the collapsing lemma is a sort of converse construction:

Proposition A. *Every closed four-arc PC curve can be obtained by starting with one based on a collapsed quadrilateral and “uncollapsing” it, keeping the radii of the circles unchanged as the quadrilateral moves away from the collapsed position.*

Note that since the radii remain unchanged, so do the edge lengths and the PC curve remains closed as the quadrilateral uncollapses. The proposition is an immediate consequence of the following lemma:

Collapsing Lemma. *Any quadrilateral satisfying (1) can be continuously collapsed so that its four vertices are collinear.*

Proof of Lemma: Let us fix $l_4 = 1$ and take $c_1 = (0, 0)$, $c_4 = (1, 0)$, $c_2 = (l \cos t, l \sin t)$ as in FIGURE 13a. The condition for c_3 to exist for this position of c_2 is $|l_2 - l_3| \leq d \leq l_2 + l_3$, where d is the distance from c_2 to c_4 . This is

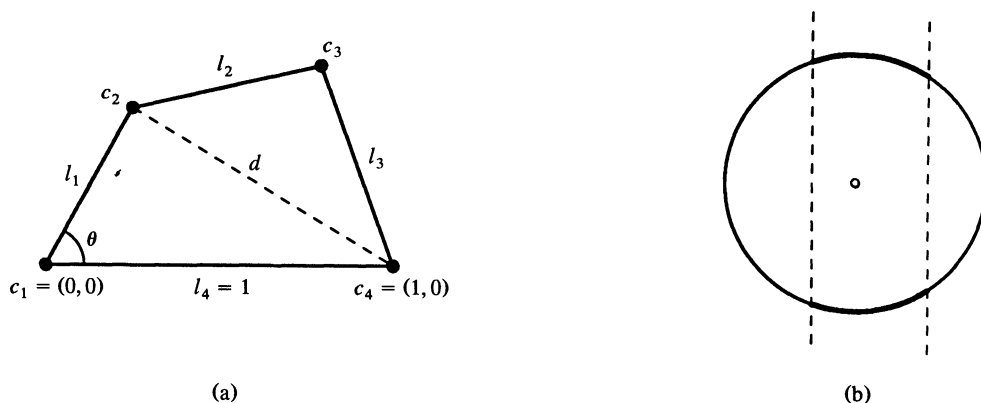


Figure 13

equivalent to

$$l_1^2 + 1 - (l_2 + l_3)^2 \leq 2l_1 \cos t \leq l_1^2 + 1 - (l_2 - l_3)^2.$$

The allowable values of t are therefore those points on the unit circle between two vertical lines, one or both of which may actually miss the circle [FIGURE 13b]. We need only check that

(i) if $l_1 \pm l_2 \pm l_3 = 1$, then $t = 0$ is an allowable value, that is, $2l_1 \leq l_1^2 + 1 - (l_2 - l_3)^2$;

(ii) if $-l_1 \pm l_2 \pm l_3 = 1$, then $t = \pi$ is an allowable value, that is, $l_1^2 + 1 - (l_2 + l_3)^2 \leq -2l_1$.

Since, in (i), $(l_1 - 1)^2 = (l_2 \pm l_3)^2$, and, in (ii), $(l_1 + 1)^2 = (l_2 \pm l_3)^2$ the results are immediate.

FIGURE 14a gives an example of a closed four-arc PC curve in which the four centers have become collinear. Despite its ordinary appearance, however, this construction is very special: in this case, the lengths satisfy $l_1 = l_3$ and $l_2 = l_4$, which implies that the quadrilateral of centers, before collapsing, was a parallelogram. As the quadrilateral unfolds, the PC curve evolves as shown in FIGURE 14b. On the other hand, when no special relation holds among the l_i , besides (1), it turns out that, when the quadrilateral has collapsed, the PC curve has become degenerate. This is the content of the following result:

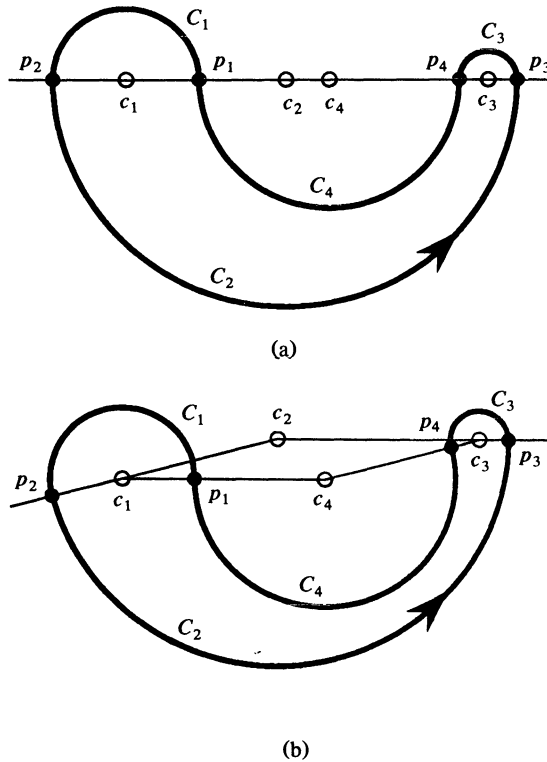


Figure 14

Proposition B. Suppose that the quadrilateral of centers collapses, with the four centers (and the four nodes) along a line. Suppose also that the four nodes p_i are all distinct. Then the quadrilateral is a parallelogram (i.e. $l_1 = l_3$ and $l_2 = l_4$).

Remark. It is “usually” true that, in the non-parallelogram case, all four nodes coincide. More precisely, if the centers c_i are all distinct and there is no restriction on the placing of the first node p_1 , then either the quadrilateral is a parallelogram or all four nodes p_i coincide.

Proof of Proposition B: Since the centers and nodes are along a line, we can take this line to be the x -axis and describe them by their x -coordinates. Since $p_2 \neq p_1$, we must have $p_2 = 2c_1 - p_1$ since c_1 is the center of the segment from p_1 to p_2 . Similarly $p_3 = 2c_2 - 2c_1 + p_1$, and $p_4 = 2c_3 - 2c_2 + 2c_1 - p_1$. Going one more step brings us back to p_1 . This gives $c_1 + c_3 = c_2 + c_4$, which implies both $l_1 = l_3$ (i.e., $|c_4 - c_1| = |c_3 - c_2|$) and $l_2 = l_4$ (i.e., $|c_1 - c_2| = |c_4 - c_3|$). The remark is proved by examining all possibilities for p_2, p_3, p_4 , given an initial p_1 .

When two consecutive nodes of a PC curve coincide, we can take the arc joining them either as a complete circle or as a mere point. FIGURE 15a–c shows a PC curve growing out of a collapsed polygon of centers where all but one of the arcs is taken as a point.

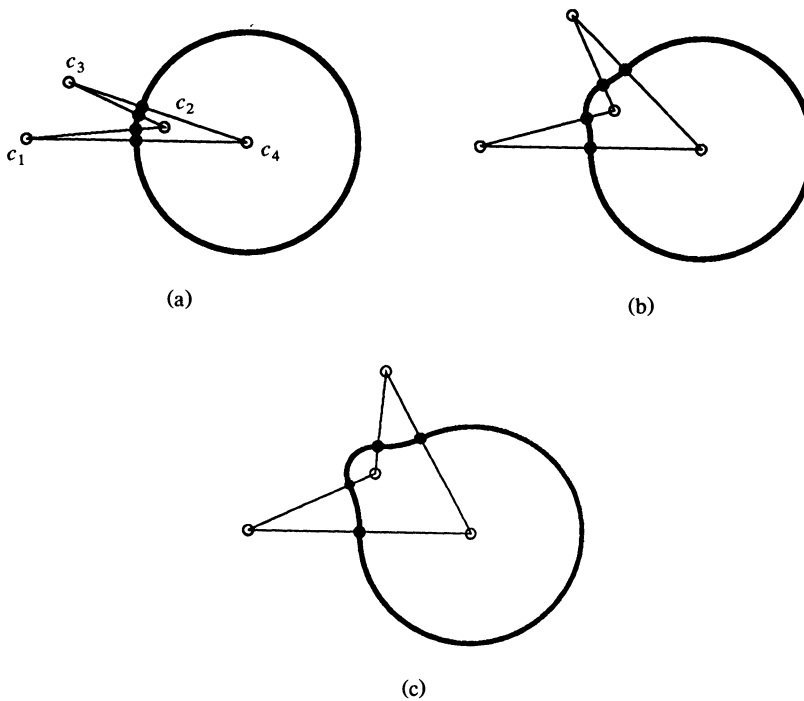


Figure 15

CONCLUDING REMARKS. Many of the topics we have introduced in this paper can be taken much further. Here we mention some natural extensions.

Closed PC curves with n arcs and a given evolute polygon fall into two classes, depending on whether the number of cusps is even or odd (smooth PC curves have zero cusps, an even number).

If the number of cusps is even, there is always a relation between the side-lengths of the polygon of the form

$$l_n = \pm l_1 \pm l_2 \pm \dots \pm l_{n-1} \quad (2)$$

analogous to (1) above. Furthermore, in this case, the radii of the PC curve can be varied to give a family of parallel PC curves. Examples are given above in FIGURES 1, 2b, 4, and 11.

If the number of cusps is odd, there is no restriction on the sides. There is a *unique* closed curve with a given evolute polygon, and the radii cannot be varied. Examples are the three-arc PC curves in FIGURES 5 and 6. Note that three-arc PC curves always have an odd number of cusps.

The extension of Proposition A and the Collapsing Lemma to PC curves with more than four arcs is straightforward but more complicated. We know of no easy proof that a polygon satisfying (2) for some choice of signs necessarily collapses continuously to a position where all the nodes are collinear. It would appear that this should be easier to achieve as n becomes larger, since the polygon becomes “floppier” as it has more degrees of freedom.

Finally we mention one remarkable example of a symmetry set of a PC curve. The four-arc PC curve C in FIGURE 16a has a *biosculating circle* S , i.e. S is an osculating circle at two points p and q . By definition, the center of S is part of the symmetry set of C , but it is an isolated point of the symmetry set since there are no circles near S that are tangent to C at two points. If we perturb the curve C by moving the node at p slightly counter-clockwise round C_1 , and adjusting C_4 and C_3 accordingly, a family of bitangent circles appears to grow out from S . An enlarged picture of the locus of centers of curvature of these bitangent circles is shown in FIGURE 16b. If we move p the other way, all of these bitangent circles come together and disappear. In the study of symmetry sets of one-parameter families of plane curves, this transition is called a *moth*. In [G-B] and [Br-G] there are extensive discussions of such transition phenomena for general smooth curves. Although many of the transitions for symmetry sets of general smooth curves already appear in the study of plane polygons, as in [Ba-G1], not all of them do, and part of our motivation for studying PC curves was an attempt to find an elementary class of curves for which all general phenomena were already present.

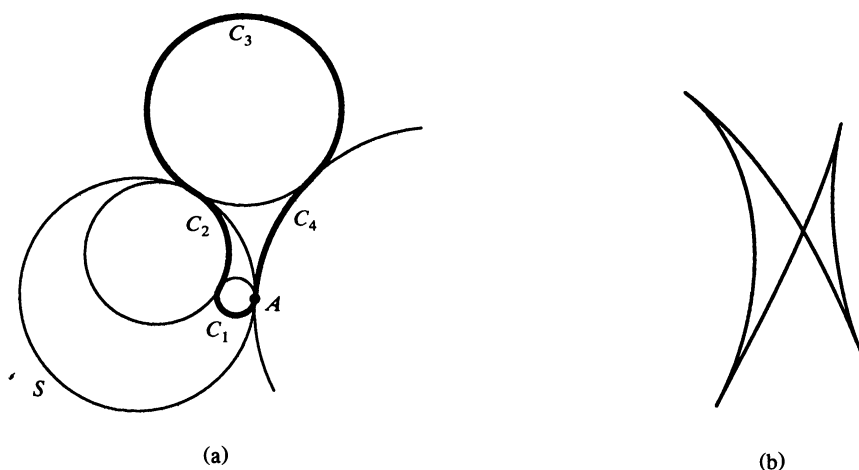


Figure 16

A full discussion of the symmetry sets of PC curves appears in [Ba-G2]. Many of the notions in this paper generalize to PC curves in space and in higher dimensions, and we intend to pursue these ideas in a subsequent paper.

We would like to thank Davide Cervone for assistance in producing the computer-generated illustrations for this paper.

REFERENCES

- [A] Artmann, Benno "The Cloisters of Hauterive" *Math. Intelligencer* 13:2 (1991) 44–49.
- [Ba-G1] Banchoff, T. and Giblin, P. "Global Theorems for Symmetry Sets of Smooth Curves and Polygons in the Plane" *Proc. Royal Soc. of Edinburgh* 106A (1987) 221–231.
- [Ba-G2] Banchoff, T. and Giblin, P. "Symmetry Sets of Piecewise Circular Curves" *Proc. Royal Soc. of Edinburgh* 123A (1993) 1135–1149.
- [Be] Besicovitch, A. S. "Variants of a Classical Isoperimetric Problem" *Quart. J. Math.* (2), 3 (1952) 42–9.
- [Br-G] Bruce, J. W. and Giblin, P. "Growth, Motion and 1-Parameter Families of Symmetry Sets" *Proc. Royal Soc. of Edinburgh* 104A (1986) 179–204.
- [Ga] Gaba, M. G. "On a Generalization of the Arbelos" *Amer. Math. Monthly* 47 (1940) 19–24.
- [G-Br] Giblin, P. and Brassett, A. "Local Symmetry of Plane Curves" *Amer. Math. Monthly* 92:10 (1985) 689–707.
- [G-N] Gibson, C. G. and Newstead, P. E. "On the Geometry of the Planar 4-Bar Mechanism" *Acta Applicandae Mathematicae* 7 (1986) 113–135.
- [H] Hood, Rodney "A Chain of Circles" *The Mathematics Teacher* (1961) 134–137.
- [M-P] Marciniak, K. and Putz, B. "Approximation of Spirals by Piecewise Circular Curves of Fewest Circular Arc Segments" *Computer Aided Design*, Vol. 16, No. 2 (1984) 87–90.
- [M-N] Martin, R. R. and Nutbourne, A. W. "Differential Geometry Applied to Curve and Surface Design" Vol. 1 (1988) Foundation Ellis Horwood.
- [O] Osserman, R. "The Four-or-More Vertex Theorem" *Amer. Math. Monthly* 92 (1985) 332–337.
- [R-R] Rossignac, J. R. and Requicha, A. A. G. Piecewise-Circular Curves for Geometric Modeling, *IBM Journal of Research and Development* (1987) 296–313.
- [S] Sabin, M. "The Use of Piecewise Forms in the Numerical Representation of Shape" Report no. 60, Computer and Automation Institute, Hungarian Academy of Science, Budapest (1977).

*Department of Mathematics
Brown University
Providence, RI 02912*

*Department of Pure Mathematics
University of Liverpool
Liverpool L69 3BX
England*

Pascal's Theorem

The proof of Pascal's Theorem mentioned in Professor van Yzeren's article (MONTHLY 100, pp. 930–931) is not my own but the proof I learned in 11th grade Descriptive Geometry and the Mathematisches-Naturwissenschaftliches Gymnasium of Basel, Switzerland. When I wrote the book I therefore assumed that the proof was part of everybody's general mathematical education. I am quite sure that this proof was absorbed by Swiss Type C (science, A is classical languages, B modern languages) students for at least 50 years. It also appears in what was the standard Swiss high school text of Descriptive Geometry (Flükiger). I did check in Italian and German D.G. texts; the Italians do not have Pascal's theorem and a German University text does not prove it. Unfortunately, I do not have Austrian high school D.G. texts but I assume that at least pre-World War I Austrian texts did present a similar proof. To find out whether Dr. van Yzeren's proof was known somewhere one would have to comb through the school literature of the few countries that did require descriptive geometry in their high school curriculum.

Unfortunately, New Math has succeeded in destroying much of European education almost as much as it did American education.

*H. Guggenheimer
P.O. Box 401
West Hempstead, NY 11552*

The Two Envelope Paradox

Elliot Linzer

In a question and answer column in *Parade* magazine [1], a reader asks:

I am asked to select one of two envelopes and told that one contains twice as much money as the other. I find \$100 in the envelope that I select. Should I switch to the other one to improve my worldly gains?

The column's author answers that while it appears that switching is a good idea—\$100 buys even odds at getting \$50 or \$200, which will average \$125—it “actually makes no difference at all” whether or not you switch. That answer is not quite correct. Indeed, the question seems to present a paradox: while it cannot always pay to switch, switching seems to increase the average take by 25%. This apparent paradox is explained in this note.

To be somewhat more specific, let us assume that a “host” chooses a random positive number and puts that many dollars in one envelope, say envelope *A*, and twice that many dollars in envelope *B*. Denote by p_x the probability that x is the number the host chose at random. (For simplicity, I am assuming a discrete distribution.) We choose one envelope (which will be *A* with probability $1/2$) and look inside. Denote by q_x the probability that the chosen envelope is envelope *A* given that we see x dollars in that envelope.

That probability is clearly the same as the probability that the host's random number was x given that it was either x or $x/2$. Therefore,

$$q_x = \frac{p_x}{p_x + p_{x/2}}.$$

If the chosen envelope contains x dollars, the expected value of the amount of money in the other envelope is $2xq_x + 0.5x(1 - q_x)$, which is larger than x when $p_x > 0.5p_{x/2}$. Thus the answer to the above question is: it depends on p_{100} and p_{50} ; because those parameters are not given, the question cannot be answered.

It must be the case that the average return from always switching is the same as the average return from always staying. Consider the following. Suppose $p_{50} = p_{100} = 1/2$, and $p_x = 0$ otherwise. There will be \$50 in the chosen envelope only if the host chooses \$50 and we pick envelope *A*. Thus with probability 0.25 we will find \$50 in the chosen envelope, and in that case we will get \$100 if we switch. Also with probability 0.25, we will find \$200 in the chosen envelope, and in that event we will get \$100 if we switch. We will find \$100 dollars in the chosen envelope with probability 0.5, (because that happens if the host chooses \$50 and we pick envelope *B* or if the host chooses \$100 and we pick envelope *A*), and in that event

we will average \$125 by switching. Overall, if we always switch we average $0.25 \times 100 + 0.5 \times 125 + 0.25 \times 100 = 112.5$ dollars. If we always stay we will average $0.25 \times 50 + 0.5 \times 100 + 0.25 \times 200 = 112.5$ dollars. By contrast, if we employ a smart strategy and switch if we find \$50 or \$100 but stay if we find \$200, we will average $0.25 \times 100 + 0.5 \times 125 + 0.25 \times 200 = 137.5$ dollars.

It is instructive to try to create a probability distribution for which it *always* pays to switch. If it always pays to switch, then we must have $p_x > 0.5p_{x/2}$ whenever it is possible to choose an envelope and find x dollars inside (That is, whenever $p_x > 0$ or $p_{x/2} > 0$.) Let's assume that that is the case and see what happens. Choose x_0 so that $p_{x_0} > 0$. Envelope B contains $2x_0$ dollars with probability $p_{x_0} > 0$, so it is possible to choose an envelope and find $2x_0$ dollars inside. Because we are assuming that it always pays to switch, we must have $p_{2x_0} > 0.5p_{x_0} > 0$. Envelope B contains $4x_0$ dollars with probability $p_{2x_0} > 0$, so it is possible to choose an envelope and find $4x_0$ dollars inside; therefore, $p_{4x_0} > 0.5p_{2x_0} > 0.25p_{x_0} > 0$. Continuing in this way, we see that

$$p_{2^i x_0} > 2^{-i} p_{x_0}.$$

The value of the random number chosen by the host is

$$\sum_x x p_x \geq \sum_{i=0}^{\infty} 2^i x_0 p_{2^i x_0} > \sum_{i=0}^{\infty} x_0 p_{x_0} = \infty.$$

Thus we can (seemingly) always increase the average take by switching only if the average amount of money that the host puts in envelope A —and, therefore, the average that we will get by switching or staying—is infinite.

As an example of when it can seem to always pay to switch, consider the following distribution. Let

$$p_{2^i} = \frac{1}{3} \left(\frac{2}{3} \right)^i$$

if i is a non-negative integer and $p_x = 0$ if x is not a power of 2. For positive integers i the probability of having chosen envelope A given that the chosen envelope has 2^i dollars in it is $q_{2^i} = 0.4$. If, for a positive integer i , we find 2^i dollars in the chosen envelope, the average amount of money that we will find in the other envelope is $2 \times 2^i \times 0.4 + 0.5 \times 2^i \times 0.6 = 1.1 \times 2^i$ dollars. The only other possible dollar amount that we can find in the chosen envelope is \$1, and in that case we will get \$2 by switching envelopes. We therefore seem to always increase the average return by switching. However, if we calculate the average return we get by always switching it will be the same as the average return we get by always staying; in both cases the average return is infinite.

It is interesting that the explanation given is in accordance with the intuitive answer to the question of whether or not to switch envelopes. If most people were faced with the choice of switching envelopes or keeping what they saw in the first envelope, they would switch if they saw a small amount but keep a large amount. This is due not only to the fact that if someone is assured he will get a very large sum of money he would not be very interested in risking half of it for larger gains (even if the odds were in his favor), but because he knows that the host only has a finite amount of money (and is probably not planning and giving too much away). So in one sense the answer may be that you've got to guess on the probability distribution that the host is using and decide accordingly.

1. M. vos Savant, Ask Marilyn, *Parade*, (September 1992) 20.

Room J1-H18
 IBM Research
 P.O. Box 704
 Yorktown Heights, NY 10598
 ELLIOTL@WATSON.IBM.COM

Disorderly Currencies

The article "Orderly Currencies" (Jan. 94, pp. 36–38) has a lovely topic and a charming style. Unfortunately, the one theorem is wrong.

A coinage system is orderly if to give change with the fewest coins it suffices to use a greedy algorithm—as many of the largest coin as possible, then of the next largest, etc. The article claims that to be orderly it is necessary and sufficient for the denominations $1 = d_1 < d_2 < \dots < d_n$ to satisfy

$$d_j \geq 2d_{j-1} - d_{j-2}, \quad 3 \leq j \leq n. \quad (1)$$

Consider a coinage with denominations 1, 6, 11, 17. These meet (1), yet the greedy algorithm on the amount 22 gives one 17 coin and five 1's, whereas only two 11's are needed. So (1) is not sufficient.

Neither is (1) necessary. Consider the denominations 1, 5, 10, 20, 25, 40. With a little work, one can show these form an orderly currency. However, for $j = 5$, condition (1) fails, as $25 < 2 \times 20 - 10$. True, if we delete the 40 piece, then the currency is not orderly. And indeed, if we insist that every initial sequence d_1, \dots, d_j be orderly, then it turns out that (1) is necessary. But, the article does not require initial sequences to be orderly, and the proof is not trivial.

There is a substantial literature on this subject, though it's hard to look up because it's hard to guess the right key words. My assertions and examples come from

M. L. Magazine, G. L. Nemhauser and L. E. Trotter Jr., *When the greedy solution solves a class of knapsack problems*. Operations Research 23 (1975) 207–217.

See Theorem 1 and its corollaries. These authors actually discuss a more general problem, which in the terminology of the *Monthly* article could be called coin changing with minimum mass.

An alternative exposition of the result of Magazine *et al.* appears in Section 6.1 of "Combinatorial Algorithms" by T. C. Hu (Addison-Wesley, 1982). Further references with brief annotations appear at the end of the chapter.

As far as I know, a good necessary and sufficient condition for orderly coinage has never been found. The article above gives a decent necessary and sufficient condition if each initial sequence of the coinage must be orderly.

Actually, before the days of electronic cash registers, which tell the clerk exactly how much change to give, clerks did *not* use the greedy algorithm, because they counted *up* to make change. That is, if you paid \$1 for an 83¢ item, the clerk counted 84 (as he gave out a penny), 85, 90 a dollar. The advantage for clerks of counting up was that they didn't need to know how to subtract! For us, though, the interesting point is that clerks were not using the greedy algorithm. Yet they (almost) always gave out the fewest coins. So what algorithm *were* clerks using, and why was it optimal? Readers might like to ponder this. I heard this question raised by Peyton Young, in December 1975 at Princeton, at the end of a seminar talk by Les Trotter about his paper. In correspondence with Trotter (unpublished) I gave an answer to this question that seems satisfactory for real-world currencies.

Stephen B. Maurer
 Department of Mathematics & Statistics
 Swarthmore College
 Swarthmore PA 19081-1397
 smaurer1@cc.swarthmore.edu

Fourier Series of Polygons

Alain Robert

It is generally admitted (among mathematicians...) that the complex form of Fourier series is the easiest to discuss. It has two main advantages

- economical on the notational side

$$\sum_{-\infty}^{\infty} c_k e^{ikt} \text{ instead of } a_0 + \sum_{k \geq 1} (a_k \cos kt + b_k \sin kt),$$

- straightforward for the discussion of absolute convergence

$$|c_k e^{ikt}| = |c_k|.$$

For the visualization of periodic functions however, one usually tends to separate real and imaginary parts, and draw separately the graphs of these two periodic functions.

But a complex Fourier series $\sum_{k \in \mathbb{Z}} c_k e^{ikt}$ represents a 2π -periodic map

$$f: \mathbb{R} \rightarrow \mathbb{C}$$

and at least when it is continuous, can be viewed as a *closed parametrized curve* $t \mapsto f(t) \in \mathbb{C}$ *in the complex plane*. In general, this closed curve will have multiple points (namely, f is not one-to-one). It is our purpose to illustrate this point of view. In particular, we intend to show that for $n \geq 2$

$$f_n(t) = \sum_{k \equiv 1(n)} e^{ikt} / k^2$$

is the Fourier series of a regular n -gon in the complex plane.

In the first picture, the Fourier series of the pentagon is illustrated: partial sums $\sum_{k \equiv 1(5), |k| \leq 5n+1} \dots$ are plotted for $n = 1, 2, 3, 4$ and 8 .

1. THE FUNDAMENTAL FOURIER SERIES AND THE MAIN RESULT. Let $\mathcal{C}(\mathbb{R}/2\pi\mathbb{Z})$ be the space of continuous 2π -periodic functions $f: \mathbb{R} \rightarrow \mathbb{C}$. We shall say that such a function f is a *polygon* if there is a finite subdivision

$$0 \leq t_0 < \dots < t_j < t_{j+1} \dots < t_{n-1} < t_n = t_0 + 2\pi$$

of $[0, 2\pi[$ such that f is *affine linear* in each subinterval $[t_j, t_{j+1}[$. In this case, the image of f is a polygonal line in \mathbb{C} with vertices $s_j = f(t_j)$, and f is a parametrization with constant speed on each side. The set of polygons is obviously a subspace of $\mathcal{C}(\mathbb{R}/2\pi\mathbb{Z})$.

Let us now determine the Fourier series of a polygon f . With the previous notations, we compute the Fourier coefficients $c_k(f)$ given by

$$2\pi c_k(f) = \int_0^{2\pi} f(t) e^{-ikt} dt = \sum_{0 \leq j < n} \int_{t_j}^{t_{j+1}} f(t) e^{-ikt} dt,$$

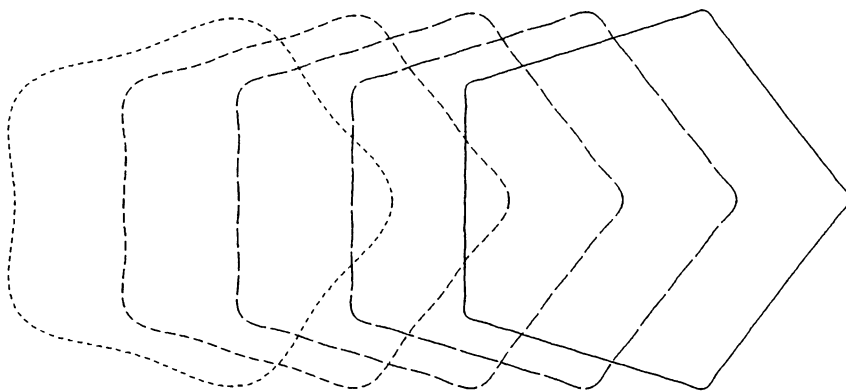


Figure 1

by integration by parts when $k \neq 0$. For a typical term

$$\int_{t_j}^{t_{j+1}} f(t) e^{-ikt} dt = \left[f(t) \frac{e^{-ikt}}{-ik} \right]_{t_j}^{t_{j+1}} - \int_{t_j}^{t_{j+1}} v_j \cdot \frac{e^{-ikt}}{-ik} dt.$$

In the sum over j , the integrated terms cancel out two by two (f is continuous and periodic) and we obtain

$$\begin{aligned} 2\pi c_k(f) &= \sum_j \int_{t_j}^{t_{j+1}} v_j \cdot \frac{e^{-ikt}}{ik} dt = \frac{1}{ik} \sum_j v_j \int_{t_j}^{t_{j+1}} e^{-ikt} dt \\ &= k^{-2} \sum_j v_j [e^{-ikt}]_{t_j}^{t_{j+1}} = k^{-2} \sum_j (v_{j-1} - v_j) e^{-ikt_j}. \end{aligned}$$

With a velocity jump $\sigma_j = v_j - v_{j-1}$ at time t_j , we can write the last relation as

$$c_k(f) = - \left(\sum_j \sigma_j e^{-ikt_j} \right) / (2\pi k^2).$$

In spite of its appearance, this is a very simple expression indeed. Recall that a *translate* τf of a function $f \in \mathcal{C}(\mathbb{R}/2\pi\mathbb{Z})$ is a function of the form

$$(\tau f)(t) = f(t - a).$$

As the integral formula shows, the Fourier coefficients of such a translate are $c_k(\tau f) = e^{-ika} c_k(f)$. We see now that a polygon f is a linear combination of translates of the basic Fourier series

$$f_1(t) = \sum_{k \neq 0} e^{ikt} / k^2$$

and a constant term $c_0 = c_0(f)$:

$$f \quad \text{and} \quad c_0 + \sum_j \alpha_j f_1(t - t_j)$$

have the same Fourier coefficients when $\alpha_j = -\sigma_j/(2\pi)$.

Conversely, if we consider a linear combination $\sum_j \alpha_j \tau_{t_j} f_1$ of translates of f_1 where $\sum_j \alpha_j = 0$, we can construct a parametrized polygon in \mathbb{C} with this Fourier series. We can indeed compute the $\sigma_j = -2\pi \alpha_j$ and the t_j , hence we can compute

the velocities v_j starting from v_0 as follows

$$v_1 = v_0 + \sigma_1, \quad v_2 = v_0 + (\sigma_1 + \sigma_2), \text{ etc.}$$

We have $v_n = v_0 + (\sigma_1 + \cdots + \sigma_n) = v_0$ by assumption. Moreover, the polygonal line will be closed when

$$(t_1 - t_0)v_0 + \cdots + (t_n - t_{n-1})v_{n-1} = 0.$$

But $t_n = 2\pi + t_0$ and reordering terms, we get

$$2\pi v_{n-1} = t_1(v_1 - v_0) + t_2(v_2 - v_1) + \cdots = \sum_{0 \leq j < n} t_j \sigma_j$$

which fixes the value of $v_0 = v_n = v_{n-1} + \sigma_n = v_{n-1} + \sigma_0 = \sigma_0 + \sum_{0 \leq j < n} t_j \sigma_j / 2\pi$. Hence the velocities are uniquely determined in all subintervals $[t_j, t_{j+1}[$ and the polygon itself is determined up to a translation. Without loss of generality, we can now restrict ourselves to *centered polygons*, namely those in $\mathcal{C}_0 = \{f \in \mathcal{C}(\mathbb{R}/2\pi\mathbb{Z}) : c_0(f) = 0\}$.

Theorem. *The space of centered polygons V consists precisely of the linear combinations $\sum \alpha_j \cdot \tau_j(f_1)$ of translates of*

$$f_1(t) = \sum_{k \neq 0} e^{ikt} / k^2$$

where $\sum \alpha_j = 0$.

As we have just seen, the study of Fourier series of polygons can be based on the basic expansion

$$f_1(t) = \sum_{k \neq 0} e^{ikt} / k^2 = 2 \sum_{k \geq 1} \cos kt / k^2.$$

Its graph is easily pictured. Since the series of derivatives converges uniformly on compact subsets of $]0, 2\pi[$ by Abel's criterium, it is legitimate to write

$$f'_1(t) = \sum_{k \neq 0} i \cdot e^{ikt} / k = -2 \sum_{k \geq 1} \sin kt / k \quad \text{for } 0 < t < 2\pi.$$

As is known, this is the Fourier series of the 2π -periodic function

$$t - \pi \quad \text{for } 0 < t < 2\pi.$$

From this we infer

$$f_1(t) = t^2/2 - \pi t + f_1(0) \quad \text{for } 0 < t < 2\pi.$$

Using the well-known $\sum_{k \geq 1} 1/k^2 = \pi^2/6$, we deduce $f_1(0) = \pi^2/3$:

$$f_1(t) = t^2/2 - \pi t + \pi^2/3 \quad \text{for } 0 < t < 2\pi.$$

The graph of f_1 is not piecewise linear (it is *not* a polygon) *but* the second derivative of a linear combination of translates $\sum \alpha_j \cdot \tau_j(f_1)$ *vanishes when* $\sum \alpha_j = 0$ because $\tau_j(f_1)'' \equiv 1$ except at finitely many points of $[0, 2\pi[$. This shows that $\sum \alpha_j \cdot \tau_j(f_1)$ is piecewise linear and gives an independent algebraic proof of the converse part of the theorem.

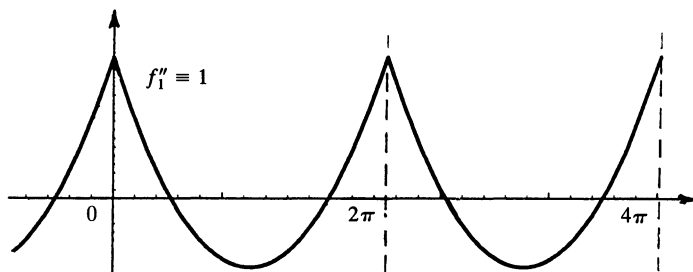


Figure 2

Corollary. *The subspace generated by the translates of f_1 is dense in \mathcal{C}_0 .*

Proof: Let $f \in \mathcal{C}_0$. Since f is uniformly continuous, for each given $\varepsilon > 0$, one can find a subdivision $\{t_j\}$ of $[0, 2\pi[$ such that

$$|f(t) - f(t_j)| \leq \varepsilon/4 \quad \text{for all } t_j \leq t \leq t_{j+1}.$$

The polygon g going through the points $f(t_j)$ is a uniform approximation of f : choosing the index j suitably, we can write

$$|f(t) - g(t)| \leq |f(t) - f(t_j)| + |f(t_j) - g(t)| \leq \varepsilon/2.$$

Now, the mean value of g will also be small

$$a = (1/2\pi) \int_0^{2\pi} g(t) dt = (1/2\pi) \int_0^{2\pi} (f(t) - g(t)) dt \text{ satisfies } |a| \leq \varepsilon/2.$$

The polygon $g_0 = g - a$ is centered, i.e. belongs to \mathcal{C}_0 and satisfies

$$\begin{aligned} |f(t) - g_0(t)| &\leq |f(t) - g(t)| + |a| \leq \varepsilon, \\ \|f - g_0\| &= \text{Sup } |f(t) - g_0(t)| \leq \varepsilon. \end{aligned}$$

2. FOURIER SERIES OF REGULAR N -GONS. We say that a continuous 2π -periodic function $f: \mathbb{R} \rightarrow \mathbb{C}$ has a symmetry of order $n \geq 2$ when there is an n th-root of unity ζ such that

$$f(t + 2\pi/n) = \zeta \cdot f(t).$$

We shall only be interested in the case $\zeta = e^{2i\pi/n}$ in this section (in Sec. 4 below, we shall give examples of the general case).

Theorem. *Let $f \in \mathcal{C}(\mathbb{R}/2\pi\mathbb{Z})$ be a continuous, 2π -periodic function presenting a symmetry of order $n \geq 2$ in the sense $f(t + 2\pi/n) = e^{2\pi i/n} f(t)$ ($t \in \mathbb{R}$). Then the Fourier sequence of f satisfies*

$$\begin{aligned} c_k(f) &= 0 \quad \text{if } k - 1 \text{ is not a multiple of } n, \\ c_k(f) &= (n/2\pi) \int_0^{2\pi/n} f(t) e^{-ikt} dt \quad \text{if } k - 1 \text{ is a multiple of } n. \end{aligned}$$

Proof: The n th order symmetry simply means that the function

$$t \mapsto e^{-it} \cdot f(t)$$

is periodic of period $2\pi/n$. It can be expanded in a Fourier series according to the

system $(e^{imnt})_{m \in \mathbb{Z}}$ and

$$f(t) = e^{it} \sum_{m \in \mathbb{Z}} c'_m e^{imnt} = \sum_{m \in \mathbb{Z}} c'_m e^{i(mn+1)t} = \sum_{k \equiv 1(n)} c_k e^{ikt}.$$

In particular, $f \in \mathcal{C}_0$. Conversely, any exponential e^{ikt} where $k \equiv 1 \pmod n$ has the n -th order symmetry property required.

In other words, the basic exponentials $(e^{ikt})_{k \equiv 1 \pmod n}$ constitute a Hilbert basis for the space of L^2 -functions with symmetry of order n .

We consider now more particularly *the case of regular polygons* in \mathbb{C} , with vertices at the n th roots s_j of 1. Put $\zeta = \zeta_n = e^{2\pi i/n}$ so that

$$s_j = e^{2\pi i j/n} = \zeta_n^j \quad (0 \leq j < n).$$

Since $\sum_{0 \leq j < n} \zeta^j = 0$, the linear combination

$$\sum_{0 \leq j < n} \zeta^j f_1(t - j \cdot 2\pi/n)$$

represents a polygon. More precisely,

$$\begin{aligned} \sum_{0 \leq j < n} \zeta^j f_1(t - j \cdot 2\pi/n) &= \sum_{0 \leq j < n} \sum_{k \neq 0} \zeta^j e^{ik(t - j \cdot 2\pi/n)} / k^2 \\ &= \sum_{k \neq 0} \sum_{0 \leq j < n} \zeta^{j(1-k)} e^{ikt} / k^2 = n \sum_{k \equiv 1(n)} e^{ikt} / k^2 = n \cdot f_n(t). \end{aligned}$$

This proves that f_n represents a polygon (constant speed on each side) with an n -th degree symmetry and only n vertices (linear combination of n translates of f). The symmetry

$$f(t + 2\pi/n) = \zeta \cdot f(t)$$

leads to

$$f'(t + 2\pi/n) = \zeta \cdot f'(t) \quad \text{and} \quad |f'(t + 2\pi/n)| = |f'(t)|$$

and since we know a priori that the velocity f'_n has only n jumps, it proves that f_n is a constant speed parametrization.

We shall now use the Parseval formula which we recall.

If f and $g \in L^2(0, 2\pi)$ have respective Fourier coefficients $a_k = c_k(f)$, $b_k = c_k(g)$, then

$$\sum \bar{a}_k b_k = (f|g) = \frac{1}{2\pi} \int_0^{2\pi} \overline{f(t)} \cdot g(t) dt.$$

In particular, for square summable 2π -periodic functions f

$$\sum |c_k(f)|^2 = \|f\|_2^2 = \frac{1}{2\pi} \int_0^{2\pi} |f(t)|^2 dt.$$

Using this identity for f'_n having Fourier coefficients

$$c_k(f'_n) = i/k \quad \text{if } k \equiv 1(n) \text{ and } 0 \text{ otherwise}$$

we deduce

$$f_n(0) = \sum_{k \equiv 1(n)} 1/k^2 = \|f'_n\|_2^2 = v_n^2.$$

The last equality holds since we know that $|f'_n| = v_n$ is constant. This speed can easily be evaluated by geometric considerations

$$2\pi v_n = L_n = 2nf_n(0) \cdot \sin \pi/n$$

hence

$$v_n = f_n(0) \cdot \frac{\sin \pi/n}{\pi/n}.$$

We have obtained

$$f_n(0) = v_n^2 = f_n^2(0) \cdot \left(\frac{\sin \pi/n}{\pi/n} \right)^2$$

from which we deduce

$$f_n(0) = \left(\frac{\pi/n}{\sin \pi/n} \right)^2.$$

We have proved the following result.

Theorem. *The Fourier series of the regular n -gon having the $s_j = e^{2ij\pi/n}$ as vertices—uniformly parametrized—is*

$$C_n f_n(t) = C_n \sum_{k \equiv 1 \pmod n} e^{ikt}/k^2 = C_n \sum_{l \in \mathbb{Z}} e^{i(1+ln)t}/(1+ln)^2$$

with normalization constant $C_n = \sin^2(\pi/n)/(\pi/n)^2$.

Corollary. *We have*

$$\sum_{l \in \mathbb{Z}} 1/(1+ln)^2 = \left(\frac{\pi/n}{\sin \pi/n} \right)^2.$$

In particular for $n = 2$, we infer

$$\sum_{k \text{ odd} \geq 1} 1/k^2 = (1/2) \sum_{l \in \mathbb{Z}} 1/(1+2l)^2 = \pi^2/8.$$

Let us observe at this point that $C_n < 1$ tends monotonously to 1 for $n \rightarrow \infty$ and $f_n \rightarrow f = e^{it}$ (uniform parametrization of the circle) uniformly for $n \rightarrow \infty$.

3. QUADRATIC SPACE V . The vector space V carries a quite natural quadratic form. Here it is.

Definition. For $f \in V$, we define $Q(f) = i\pi(f'|f) = \pi((1/i)f'|f)$.

Theorem. *The function Q takes real values only and defines a non-degenerate real quadratic form on V . Moreover*

1. $Q(f) = \pi \sum k |c_k(f)|^2$,
2. When $f = \partial\Pi$ is the (positively oriented) boundary of a polygonal piece $\Pi \subset \mathbb{C}$, $Q(f) = S = \text{Area}(\Pi)$.

Proof: 1 results from the Parseval formula since the Fourier coefficients of $(1/i)f'$ are the $kc_k(f)$. To prove 2, we use Stokes' theorem

$$\begin{aligned} Q(f) &= i\pi(f'|f) = i\pi/(2\pi) \int_0^{2\pi} \overline{f'(t)} f(t) dt = i/2 \oint_{\partial\Pi} z d\bar{z} \\ &= (i/2) \int \int_{\Pi} dz \wedge d\bar{z} = (i/2) \int \int_{\Pi} (dx + i dy) \wedge (dx - i dy) \\ &= \int \int_{\Pi} dx \wedge dy = S. \end{aligned}$$

Corollary. $\sum_{k \equiv 1 \pmod n} 1/k^3 = Q(f_n)/\pi = ((\pi/n)/\sin \pi/n))^3 \cos \pi/n$.

Proof: The area of the regular n -gon inscribed in a circle of radius $f_n(0)$ is indeed $S = n \cdot f_n^2(0) \cdot \cos \pi/n \cdot \sin \pi/n$ as is easily seen.

4. EXAMPLES. Fix the integer $n > 1$ and choose another integer $1 \leq a < n$. Then

$$\sum_{k \equiv a \pmod n} e^{ikt}/k^2 = (1/n) \sum_{0 \leq j < n} \zeta^{aj} f_1(t - j \cdot 2\pi/n)$$

is a star-shaped polygon with $n/(n, a)$ sides (with n sides when a is relatively prime to n). The first side links $s_0 = 1$ to $s_1 = \zeta^a$, the second links ζ^a to $\zeta^{2a} \dots$. This parametrization f has an n th-order symmetry of the type

$$f(t + 2\pi/n) = \zeta^a \cdot f(t).$$

Here is an example treated with MATHEMATICA™.

Another type of stars (without double point) is obtained as follows. For two constants A and B consider

$$A \sum_{k \equiv 1 \pmod{2n}} e^{ikt}/k^2 + B \sum_{k \equiv n+1 \pmod{2n}} e^{ikt}/k^2.$$

```
f[t_]:=Sum[Exp[(2+5k) I t]/(2+5k)^2,{k,-4,4}]
u[t_]:=Re[f[t]]
v[t_]:=Im[f[t]]
ParametricPlot[{u[t],v[t]},{t,1,1+2Pi},AspectRatio->Automatic]
```

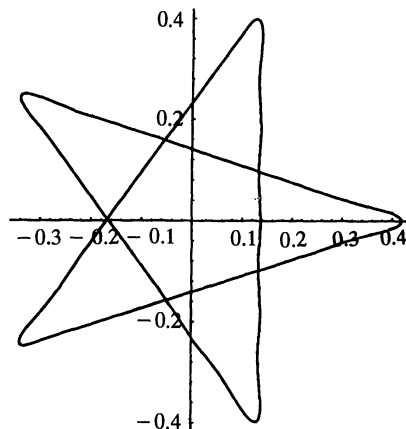


Figure 3

This is a star with n vertices. The choice

$$A = n - \sqrt{n}, \quad B = n + \sqrt{n}$$

leads to the classical stars where the sides $[s_j, s_{j+1}]$ and $[s_{j+3}, s_{j+4}]$ are on the same line when j is even. We give two more examples with MATHEMATICA™.

```
f[t_]:=Sum[Exp[(1+20k) I t]/(1+20k)^2,{k,-2,2}]
g[t_]:=Sum[Exp[(11+20k) I t]/(11+20k)^2,{k,-2,2}]
h[t_]:=f[t]-6g[t]
u[t_]:=Re[h[t]]
v[t_]:=Im[h[t]]
ParametricPlot[{u[t],v[t]},{t,1,1+2 Pi},AspectRatio->Automatic]
```

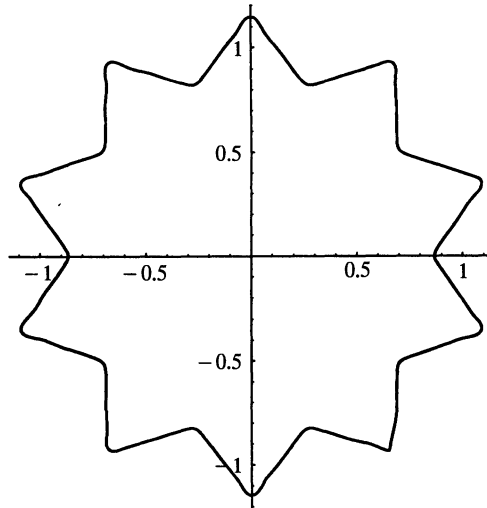


Figure 4

```
f[t_]:=Sum[Exp[(1+16k) I t]/(1+16k)^2,{k,-4,4}]
g[t_]:=Sum[Exp[(9+16k) I t]/(9+16k)^2,{k,-4,4}]
h[t_]:=f[t]+3g[t]
u[t_]:=Re[h[t]]
v[t_]:=Im[h[t]]
ParametricPlot[{u[t],v[t]},{t,1,1+2 Pi},AspectRatio->Automatic]
```

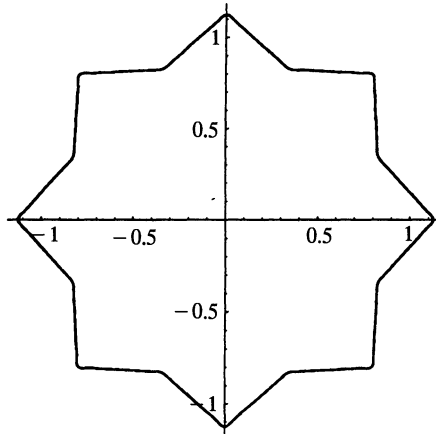


Figure 5

5. CONCLUDING COMMENTS. The interested reader may continue by programming the Fourier series of a cross (degenerate star with four sides) and check that $Q(f) = 0$ for these functions.

The formula $Q(f) = \pi \sum k |c_k(f)|^2$ is reminiscent of the Sylvester decomposition of the quadratic form Q . Indeed, Q is defined on the *Sobolev space*

$$H^{1/2} = \left\{ (c_k)_{k \in \mathbb{Z}} : \sum |k| |c_k|^2 < \infty \right\}$$

which decomposes as a direct sum $H^{1/2} = H_+ \oplus \mathbb{C} \oplus H_-$ with

$$H_+ = \{(c_k) : c_k = 0 \text{ for } k \leq 0\}, H_- \text{ similarly defined.}$$

With respect to this decomposition, $Q = Q_+ \oplus 0 \oplus -Q_-$, where Q_{\pm} are positive non degenerate on H_{\pm} .

In Fig. 3, $Q(f)$ counts twice the area of the portion of the star which is the inner pentagon. In general, when f is self intersecting, one could introduce its *algebraic area* as being $Q(f)$. Geometrically, in the computation of this area, each connected component of $\mathbb{C} - \text{Image}(f)$ is affected by a rational integer, the *index of f with respect to the points in this component*. Observe that the method of determining $f(0)$ also applies to stars since we still have a constant speed parametrization:

$$f(0) = \|f'\|_2^2 = v^2 = (\text{length}/2\pi)^2.$$

Several generalizations of the preceding considerations can be developed:

- \mathcal{C}^1 curves in \mathbb{C} which are piecewise quadratic,
- polygons in \mathbb{C}^n or \mathbb{R}^n (in particular in \mathbb{R}^3 !).

Question. Replacing the circle group by the rotation group, is it possible to describe in a simple way the Platonic solids by means of spherical harmonics?

Finally, let me thank R. S. Strichartz whose comments helped me to write the final version of this paper.

REFERENCES

1. R. Edwards, *Fourier Series*, Springer-Verlag GTM 64, 1979.
2. A. Robert, *Advanced Calculus for Users*, North-Holland, 1989.

Institut de Mathématiques
Université de Neuchâtel
Chantemerle 20
CH-2007 Neuchâtel SWITZERLAND
alain.robert@maths.unine.ch

The Paradox of Nontransitive Dice

Richard P. Savage, Jr.

Suppose one takes three cubes and writes a number from 1 to 18 on each face of each cube using each number once and only once. In this way one constructs some unusual dice we denote by A , B , and C . Is it possible to arrange the numbers on the dice such that if the dice are rolled, the probability that A beats B is greater than $1/2$, the probability that B beats C is greater than $1/2$, yet the probability that C beats A is also greater than $1/2$? The answer is yes. For example, put 18, 9, 8, 7, 6, and 5 on A ; 17, 16, 15, 4, 3, and 2 on B ; and 14, 13, 12, 11, 10, and 1 on C . It is easily checked that the probability that A beats B is $21/36$, that B beats C is $21/36$, and that C beats A is $25/36$. This phenomenon of “nontransitive dice” was popularized in Martin Gardner’s column in *Scientific American* [7]. He gives an example due to Bradley Efron with four dice in which A beats B , B beats C , C beats D , and D beats A each with probability $2/3$.

Nontransitive dice fall into the general category of nontransitivity paradoxes about which there has been considerable study. The best known nontransitivity paradox is the voting paradox the study of which was begun by Condorcet [4]. In the case of three candidates this is the observation that a majority of voters may prefer candidate A to candidate B , B to C , and C to A . For a history of the voting paradox see Black [1]. Other examples of nontransitivity are discussed in [2], [5], and [8].

Let us consider the following game. A casino has constructed three dice with the numbers arranged on them. A patron chooses whichever die he wishes and the house then picks one of the remaining dice (the one which beats the patron’s choice) and they roll the dice with a wager on the outcome. Assuming that the players will play to their own advantage, the goal that the casino has in arranging the numbers originally must be to make the smallest of the three probabilities as large as possible. In the following discussion the fact that a standard die has 6 faces is not important to the analysis, thus we consider the problem for n -faced dice.

Let X be a random variable taking as values the numbers on die A each with probability $1/n$. Define random variables Y and Z for dice B and C similarly. Clearly X , Y , and Z are independent. It was noted by Steinhaus and Trybula [9], [11] that if X , Y , and Z are random variables it is possible that $P(X > Y)$, $P(Y > Z)$, and $P(Z > X)$ can all exceed $1/2$, a result which is known as the Steinhaus-Trybula paradox. Further, they announced that if X , Y , and Z are independent then at least one of the probabilities is no more than $(\sqrt{5} - 1)/2$. (The golden ratio again!) Consequently we have the following result.

Theorem 1. Suppose numbers $1, 2, \dots, 3n$ are arranged on n -sided dice A , B , and C . Then at least one of the probabilities (A beats B , B beats C , and C beats A) is less than $(\sqrt{5} - 1)/2$.

A proof of this result in the present context is included in this paper.

More generally suppose we have m independent random variables (or m n -sided dice). Let p_{m-1} denote the unique solution of the system $p_1 = p_2(1 - p_1) = \cdots = p_{m-1}(1 - p_{m-2}) = 1 - p_{m-1}$ where for each i , $1/4 < p_i < 3/4$. Then Trybula [12] showed (see also [13]) that

$$\min\{P(X_1 > X_2), P(X_2 > X_3), \dots, P(X_m > X_1)\} \leq p_{m-1}.$$

Further, the sequence $\{p_i\}$ is increasing and converges to $3/4$. For example, if $m = 4$ it is easy to determine that $p_3 = 2/3$. Hence Efron's example is optimal.

Returning to the problem of three dice, let us now consider placing the numbers on the n -sided dice in a systematic way. One first puts one or more of the highest numbers on A . Then one puts the highest numbers that remain (one or more) on die B . Then one goes to die C (or perhaps back to A) and puts the highest numbers that remain on die C (or A). This continues until all the numbers are placed. Form a sequence x_1, x_2, \dots, x_{3k} by letting x_1 be the number of numbers placed on die A initially, x_2 the number on die B initially, x_3 the number on die C initially, x_4 the number on die A on the second go round and so on. Eventually the process stops. In the example already given $x_1 = 1$, $x_2 = 3$, $x_3 = 5$, $x_4 = 5$, $x_5 = 3$, and $x_6 = 1$. Note that the sequence has the following properties.

$$\sum_{i=1}^k x_{3i-2} = \sum_{i=1}^k x_{3i-1} = \sum_{i=1}^k x_{3i} = n. \quad (1)$$

Each x_i is a nonnegative integer. (2)

Also note that the probabilities in question are given by

$$\frac{1}{n^2} \sum_{i=1}^k x_{3i-2} \left(\sum_{j \geq i} x_{3j-1} \right), \quad \frac{1}{n^2} \sum_{i=1}^k x_{3i-1} \left(\sum_{j \geq i} x_{3j} \right), \quad \text{and} \quad \frac{1}{n^2} \sum_{i=1}^n x_{3i} \left(\sum_{j \geq i} x_{3j+1} \right). \quad (3)$$

Now allow the numbers in the sequence $\{x_i\}$ to be nonnegative real numbers instead of just nonnegative integers. The formulas for the probabilities are then continuous functions of the variables and therefore if we define F by letting F be the minimum of the three probabilities F is continuous. As F is defined on a compact set, F must achieve a maximum. If $x_j \neq 0$ but $x_m = 0$ whenever $m > j$ we will say that the length of the sequence is j . Also let α denote the number $(\sqrt{5} - 1)/2$.

The proof of Theorem 1 may be divided into three steps.

Step 1. We show that if $\{x_i\}$ is a sequence of length 6 or more than there is a sequence $\{z_i\}$ of the same length in which one of the terms is zero and for which each of the probabilities in (3) is at least as large as for $\{x_i\}$.

Step 2. We show that from a sequence in which one term is zero a new sequence can be constructed which is shorter and for which the minimum of the probabilities is at least as large as the minimum for the original sequence.

Step 3. Steps 1 and 2 combine to show that to maximize the probabilities we need only consider sequences of length 5 or less. This maximum is calculated in this step by showing that $x_1 = \alpha^2 n$, $x_2 = \alpha n$, $x_3 = n$, $x_4 = \alpha n$, and $x_5 = \alpha^2 n$ maximizes F .

Remark. The proof of the theorem for m n -sided dice is analogous. For example, in Step 3 the sequences to be considered are of length $2m - 1$.

Proof of Theorem 1.

Proof of Step 1: Note that by relabelling the dice if necessary we may assume that $x_1 \neq 0$ and $x_2 \neq 0$. If some $x_m = 0$ where $m \geq 3$ and m is less than the length of the sequence we may proceed to Step 2. If not, we will find a new sequence for which some term is zero. To get the desired sequence $\{z_i\}$ we will only change the first 6 terms of $\{x_i\}$. Therefore if $i \geq 7$ define $z_i = x_i$. Define $z_3 = \lambda x_3$ where $\lambda > 0$. By (1) this forces $z_6 = x_6 + (1 - \lambda)x_3$. We will choose z_4 such that the third of the probabilities in (3) is unchanged. It is easy to check that this forces $z_4 = (1/\lambda)x_4$. Then by (1) again we must have $z_1 = x_1 + (1 - (1/\lambda))x_4$. We would also like the first probability in (3) to be unchanged. A short computation shows that this forces $z_2 = \lambda x_2$ and therefore $z_5 = x_5 + (1 - \lambda)x_2$.

Now comparing the second of the probabilities for $\{z_i\}$ with that for $\{x_i\}$ we get

$$\begin{aligned} & \sum_{i=1}^k z_{3i-1} \left(\sum_{j \geq i} z_{3j} \right) - \sum_{i=1}^k x_{3i-1} \left(\sum_{j \geq i} x_{3j} \right) \\ &= (z_2 - x_2)n + z_5(n - z_3) - x_5(n - x_3) \\ &= x_3 x_5 - \lambda x_3 [(1 - \lambda)x_2 + x_5] \\ &= (1 - \lambda)x_3(x_5 - \lambda x_2) \end{aligned} \quad (*)$$

First suppose $x_2 \geq x_5$. The expression $(*)$ is then increasing in λ for $\lambda \geq 1$. Hence by choosing λ appropriately we can form a sequence of nonnegative reals with either $z_5 = 0$ or $z_6 = 0$ and for which $F(\{z_i\}) \geq F(\{x_i\})$. On the other hand, if $x_2 < x_5$ the expression $(*)$ is decreasing for $\lambda \leq 1$. Choose λ such that $z_1 = 0$. Again we have $F(\{z_i\}) \geq F(\{x_i\})$.

Proof of Step 2: If $x_2 < x_5$ so that $z_1 = 0$ it is trivial that the sequence $\{y_i\}$ defined by $y_i = z_{i+1}$ is shorter and has the same probabilities. In the other case, we have $z_m = 0$ where $m \geq 3$. If the length of the original sequence was 6 and $m = 6$ we have found our desired sequence. If not, define a new sequence $\{y_i\}$ by $y_j = z_j$ if $j < m - 2$, $y_{m-2} = z_{m-2} + z_{m+1}$, $y_{m-1} = z_{m-1} + z_{m+2}$, $y_m = z_m + z_{m+3}$, and $y_j = z_{j+3}$ if $j > m$. It can be checked that two of the three sums in the definition of F are the same for $\{z_i\}$ as for $\{y_i\}$. The sum involving $y_{m-2}y_{m-1} = (z_{m-2} + z_{m+1})(z_{m-1} + z_{m+2})$ has the additional term $z_{m+1}z_{m-1}$ so is at least as large as the sum for $\{z_i\}$. Hence in this case we get a sequence $\{y_i\}$ shorter than $\{x_i\}$ with $F(\{y_i\}) \geq F(\{x_i\})$. As an illustration of this idea let $\{x_i\}$ be a sequence with $x_3 = 0$ and form the sequence $x_1 + x_4, x_2 + x_5, x_6, x_7, \dots$. Clearly the probability that A beats B has increased while the other probabilities are unchanged.

Proof of Step 3: The three expressions under consideration are then $[x_1 n + (n - x_1)(n - x_2)]/n^2$, x_2/n , and $(n - x_1)/n$. The first expression is increasing as a function of x_1 while the third is decreasing. The first expression is decreasing as a function of x_2 while the second is increasing. Thus for a sequence which maximizes F the three expressions must be equal for if they were not we could increase the smallest (or two smaller) while decreasing the larger while preserving the order of the three expressions. Setting them equal we immediately get $x_2 = n - x_1$ and

completing the algebra we have

$$x_1 = \frac{3 - \sqrt{5}}{2}n = \alpha^2 n \quad \text{and} \quad x_2 = \alpha n.$$

The formulas for x_4 and x_5 follow immediately.

It is immediate that all three probabilities yield α for this sequence, completing the proof.

It is interesting to ask how close we can come to achieving the upper bound on the probabilities in Theorem 1. This question does not appear to have been considered in the literature, however see [6], [10], and [14] for other results concerning nontransitive dice. Of course, since α is irrational at least one of the probabilities must be strictly less than α . In terms of rational numbers, the result could be stated as saying that at least one of the probabilities does not exceed $\lfloor \alpha n^2 \rfloor / n^2$ where $\lfloor \cdot \rfloor$ denotes the greatest integer function. Recall that if F_k denotes the k th Fibonacci number then $\lim_{k \rightarrow \infty} F_k / F_{k+1} = \alpha$. The proof of Theorem 1 suggests that if n is a Fibonacci number then we should be able to come close to the upper bound. The following theorem shows that in this case it is possible to arrange the numbers so that all of the probabilities are at least $\lfloor \alpha n^2 \rfloor / n^2$. Hence, in this case we can achieve the (rational) upper bound.

Theorem 2. *If n is a Fibonacci number it is always possible to arrange the numbers $1, 2, \dots, 3n$ on three n -sided dice such that each of the probabilities is greater than $(1 - 1/n^2)\alpha$. Each of the probabilities is then at least $\lfloor \alpha n^2 \rfloor / n^2$.*

Proof: Let $n = F_k$. Let $x_1 = F_{k-2}$, $x_2 = F_{k-1}$, $x_3 = F_k$, $x_4 = F_{k-1}$, $x_5 = F_{k-2}$, and $x_i = 0$ if $i \geq 6$. The three probabilities are $(F_{k-2}F_k + F_{k-1}F_{k-2})/F_k^2$, F_{k-1}/F_k , and F_{k-1}/F_k . With the aid of the easily checked identity $F_{k-1}F_{k+1} - F_k^2 = (-1)^k$ one can check that the sequence F_{2j}/F_{2j+1} is increasing, that F_{2j-1}/F_{2j} is decreasing, and that the sequences have a common limit of α . Then $F_{k-1}/F_k < \alpha$ if k is odd and $F_{k-1}/F_k > \alpha$ if k is even. Now considering the three probabilities it is trivial that $F_{k-1}/F_k > (1 - 1/F_k^2)\alpha$ if k is even and follows from easy manipulations if k is odd. Now

$$\frac{F_{k-2}F_k + F_{k-1}F_{k-2}}{F_k^2} = \frac{F_k^2 - F_{k-1}^2}{F_k^2}$$

and

$$\frac{F_k^2 - F_{k-1}^2}{F_k^2} > \left(1 - \frac{1}{F_k^2}\right)\alpha \quad \text{if and only if} \quad \frac{F_k^2 - F_{k-1}^2}{F_k^2 - 1} > \alpha.$$

The result now follows from the identity

$$\frac{F_k^2 - F_{k-1}^2}{F_k^2 - 1} = \begin{cases} F_{k+1}/F_{k+2} & \text{if } k \text{ is even} \\ F_{k-2}/F_{k-1} & \text{if } k \text{ is odd.} \end{cases}$$

The last claim follows from noting that $\alpha - (1 - 1/n^2)\alpha = \alpha/n^2 < (1/n^2)$. Thus, interpreting the result in terms of rational numbers we see that each probability is at least $\lfloor \alpha n^2 \rfloor / n^2$.

In general, we cannot always achieve the upper bound of Theorem 1. In what follows we simplify the notation by denoting x_1, x_2, x_3 by x, y, z and we will let $x_4 = n - x$, $x_5 = n - y$, and $x_6 = n - z$. Motivated by the proofs of Theorems 1

and 2, in order for the probabilities to be close to the upper bound of Theorem 1 we would expect y to be close to αn . Try choosing y to be the integer in $(\alpha n + \delta - 1, \alpha n + \delta)$ where δ is a number between 0 and 1 which will be specified later. Then set $z = n$ and $x = n - y$. The numerators in the three probabilities are then $n^2 - y^2$, ny , and ny . Now $n^2 - y^2$ is smallest at the right hand endpoint while ny is smallest at the left. Choose δ such that $n^2 - (\alpha n + \delta)^2 = n(\alpha n + \delta - 1)$. This simplifies to $n = (2\alpha + 1)n\delta + \delta^2$ so $\delta \approx 1/(2\alpha + 1)$. It is now easily checked that the smallest of the numerators on the interval is

$$n^2 - (\alpha n + \delta)^2 = \alpha n^2 - \frac{2\alpha}{2\alpha + 1}n - \frac{1}{(2\alpha + 1)^2}.$$

Hence we have shown that it is always possible to arrange the numbers on the dice such that each of the probabilities exceeds

$$\alpha - \frac{2\alpha}{2\alpha + 1} \frac{1}{n} - \frac{1}{(2\alpha + 1)^2} \frac{1}{n^2}.$$

For 6-sided dice this construction yields $x = 2$, $y = 4$, and $z = 6$ for which the probabilities can be calculated to be $20/36$, $24/36$, and $24/36$. This is not as good as the example given at the beginning. In fact, we can obtain a stronger result if more freedom is allowed in specifying x and z .

Theorem 3. *It is always possible to arrange the numbers $1, 2, \dots, 3n$ on the dice such that each of the probabilities exceeds*

$$\alpha - \frac{2\alpha^2}{2\alpha + 1} \frac{1}{n} + \frac{\alpha^2}{2\alpha + 1} \frac{1}{n^2} - \frac{1}{2\alpha + 1} \frac{1}{n^3}.$$

Proof: Let $(\alpha n + \delta - 1, \alpha n + \delta)$ be an interval containing αn where δ is to be determined. Let y be the unique integer in this interval. Then we will choose x and z in one of the following three ways.

- (1) $x = n - y$ and $z = n$
- (2) $x = n - y - 1$ and $z = n - 1$
- (3) $x = n - y - 2$ and $z = n - 1$

The numerators in the three probabilities are then

- (1) $n^2 - y^2, ny, ny$
- (2) $n^2 - y^2 - y, ny + (n - y), ny + (n - y) - 1$
- (3) $n^2 - y^2 - 2y, ny + (n - y), ny + (2n - y) - 2$.

First set $ny = n^2 - y^2 - y$. Writing $y = \alpha n - t$ and simplifying we get $\alpha n = (2\alpha + 1)nt - t^2 + t$. Clearly this equation has a solution t_0 with $0 < t_0 < 1$. Then $\alpha n' < (2\alpha + 1)nt_0 + t_0$ and $\alpha n > (2\alpha + 1)nt_0$ from which we obtain

$$\frac{\alpha}{(2\alpha + 1) + \frac{1}{n}} < t_0 < \frac{\alpha}{2\alpha + 1}.$$

If y is in the interval

$$I_1 = \left(\alpha n - \frac{\alpha}{2\alpha + 1}, \alpha n + \delta \right)$$

choose x and z as in (1). Note that to the left of αn the smallest of these

numerators is ny and in I_1 ,

$$ny > n\left(\alpha n - \frac{\alpha}{2\alpha + 1}\right) = \alpha n^2 - \frac{\alpha n}{2\alpha + 1}$$

which exceeds the bound in the statement of the theorem.

Now set $ny + (n - y) - 1 = n^2 - y^2 - 2y$ and write $y = \alpha n - t$. This yields $(\alpha + 1)n - 1 = (2\alpha + 1)nt + t - t^2$. Again, there is a solution with $0 < t_1 < 1$. In the same way as before we find

$$\frac{\alpha + 1 - \frac{1}{n}}{(2\alpha + 1) + \frac{1}{n}} < t_1 < \frac{\alpha + 1 - \frac{1}{n}}{2\alpha + 1}.$$

If y is in the interval

$$I_2 = \left(\alpha n - \frac{\alpha + 1 - \frac{1}{n}}{2\alpha + 1}, \alpha n - \frac{\alpha}{2\alpha + 1} \right)$$

choose x and z as in (2). We see from the definition of t_0 that

$$n^2 - y^2 - y > \alpha n^2 - \frac{\alpha n}{2\alpha + 1}$$

in this interval, and in I_2

$$\begin{aligned} yn + (n - y) - 1 &> \left(\alpha n - \frac{\alpha + 1 - \frac{1}{n}}{2\alpha + 1} \right) n + \left(n - \alpha n + \frac{\alpha + 1 - \frac{1}{n}}{2\alpha + 1} \right) - 1 \\ &= \alpha n^2 - \frac{2\alpha^2}{2\alpha + 1} n + \frac{\alpha^2}{2\alpha + 1} - \frac{\frac{1}{n}}{2\alpha + 1} \end{aligned}$$

which gives the bound in the theorem.

Finally if y is in the interval

$$I_3 = \left(\alpha n + \delta - 1, \alpha n - \frac{\alpha + 1 - \frac{1}{n}}{2\alpha + 1} \right)$$

choose x and z as in (3). Now from the definition of t_1 , $n^2 - y^2 - 2y$ exceeds the bound of the theorem in I_3 .

In the part of I_1 to the right of αn the smallest numerator in (1) is $n^2 - y^2$ and $n^2 - y^2 > n^2 - (\alpha n + \delta)^2$ on I_1 . Now consider the numerator $ny + (n - y)$ from (3). This exceeds $n(\alpha n + \delta - 1) + (n - (\alpha n + \delta - 1))$ on I_3 . Define δ_0 to be the number between 0 and 1 for which these expressions are equal. Making estimates as before we find

$$\frac{\alpha - \frac{1}{n}}{2\alpha + 1} < \delta_0 < \frac{\alpha - \frac{1}{n}}{(2\alpha + 1) - \frac{1}{n}}.$$

Now define

$$\delta = \frac{\alpha - \frac{1}{n}}{2\alpha + 1}.$$

It can be routinely checked that $n^2 - y^2$ exceeds the bound in the theorem at the right hand endpoint as does $ny + (n - y)$ at the left. This completes the proof.

In the following table only the numerators are given for the various probabilities. The denominators in each case are n^2 . The upper bound in column 2 is the upper bound on the smallest of the probabilities given by Theorem 1 truncated to two places to the right of the decimal. The lower bound is the bound of Theorem 3. The values of x , y and z are those from the proof of Theorem 3. With x , y and z as given, the column "highest" gives the numerator in the highest of the three probabilities and the column "lowest" gives the smallest.

n	Upper bound	Lower bound	x	y	z	Highest	Lowest
3	5.56	4.55	0	1	2	6	5
4	9.88	8.58	1	2	3	10	9
5	15.45	13.82	2	3	5	16	15
6	22.24	20.29	1	3	5	25	21
7	30.28	27.99	2	4	6	31	29
8	39.55	36.93	3	5	8	40	39
9	50.06	47.10	3	5	8	51	48
10	61.80	58.51	4	6	10	64	60
11	74.78	71.15	4	7	11	77	72
12	88.99	85.03	4	7	11	89	88
13	104.44	100.14	5	8	13	105	104
14	121.13	116.49	5	8	13	124	117
15	139.05	134.07	6	9	15	150	135
16	158.21	152.89	6	10	16	160	156
17	178.61	172.94	6	10	16	179	176
18	200.24	194.23	7	11	18	203	198
19	223.11	216.76	6	11	18	234	217
20	247.21	240.52	7	12	19	248	244

REFERENCES

1. D. Black, *The Theory of Committees and Elections*, Cambridge: Cambridge University Press (1958).
2. C. Blyth, Some probability paradoxes in choice from among random alternatives, *J. Amer. Statist. Assoc.* 67 (1972), 366–373.
3. Chang Li-chien, On the maximin probability of cyclic random inequalities, *Scientia Sinica* 10 (1961), 499–504.
4. Marquis de Condorcet, *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Paris (1785).
5. W. W. Funkenbusch, A gaming wheel based on cyclic advantage in symbol choice, *The Gambling Papers*, Vol. XIII (1982), 68–83 University of Nevada, Reno.
6. W. W. Funkenbusch and Saari, D. G., Preferences among preferences or nested cyclic stochastic inequalities, *Congr. Numer.* 39 (1983), 419–432.
7. M. Gardner, The paradox of the nontransitive dice and the elusive principle of indifference, *Scientific American* 223 (1970), 110–114.
8. M. Gardner, On the paradoxical situations that arise from nontransitive relations, *Scientific American*, 231 (1974), 120–125.
9. H. Steinhaus and S. Trybula, On a paradox in applied probabilities, *Bull. Acad. Polon. Sci.* 7 (1959), 67–69.

10. R. L. Tenney and C. C. Foster, Non-transitive dominance, *Math. Mag.* 49 (1976), 115–120.
11. S. Trybula, On the paradox of three random variables, *Zastos. Mat.* 5 (1961), 331–332.
12. S. Trybula, On the paradox of n random variables, *Zastos. Mat.* 8 (1965), 143–154.
13. Z. Usiskin, Max-min probabilities in the voting paradox, *Ann. Math. Statist.* 35 (1964), 857–862.
14. R. Wynegar, *Strategies for the Niven dice game*, Master's thesis, Tennessee Technological University (1987).

Department of Mathematics
Box 5054
Tennessee Technological University
Cookeville, TN 38505

Hardy and Pólya

Enclosed are contradictory quotations from two well-known mathematicians. They may be of some interest to your readers.

Hardy believed that, in mathematics, the Cambridge examinations called “triposes” were nonsensical. As a demonstration, he persuaded George Pólya (who, if anything, was a master of computation and manipulation in classical analysis) to take the mathematics tripos without previous coaching. Pólya supposedly failed miserably.—Stanislaw M. Ulam, *Adventures of a Mathematician* (Scribners, 1970), page 60.

I wrote to Professor Pólya to ask him about this reference but never received a reply. He died a few years later. Then I came across the following note.

While Pólya was at Cambridge, Hardy was in the midst of his campaign to reform the mathematical Tripos and asked Pólya to take the exam unofficially. Hardy expected Pólya's poor showing would demonstrate that most of the questions on the Tripos were irrelevant to “modern continental mathematics”. Unfortunately for Hardy's plan, Pólya's performance was the best on the examination, and he would have been named Senior Wrangler if he had been a student.—Halsey Royden, “A History of Mathematics at Stanford,” *A Century of Mathematics in America*, vol. 2 (American Mathematical Society, 1989), pages 250–251.

Perhaps some of the readers of the *Monthly* could clear up this apparent discrepancy. Since Professor Halsey was Pólya's colleague, and Pólya was Pólya, I suspect the second quotation is correct. Incidentally, Hardy was fourth in the mathematical tripos of 1898.

Michael Stueben
 4651 Brentleigh Court
 Annandale, VA 22003
mstueben@tjhsst.vak12.ed.edu

Squares Expressible as Sum of Consecutive Squares

Laurent Beekmans

1. INTRODUCTION. The following problem was proposed in 1875 by Lucas [2]: *when does a square pyramid of cannon-balls contain a number of cannon-balls which is a perfect square?* This problem is clearly equivalent to solving the Diophantine equation $1^2 + \cdots + k^2 = \square$. Lucas claimed that the only solutions were $k = 1$ and $k = 24$, but this was not proved correctly until 1918 by Watson [5]. A more general problem is to determine the set S of values of k for which there exists a square equal to the sum of k consecutive squares. For example, $600 \in S$ since

$$25^2 + 26^2 + \cdots + 623^2 + 624^2 = 9010^2.$$

Little has been published about this problem but a good account of it can be found in [6], where it is shown that S is infinite and has density zero. Moreover, [6] contains a table giving all the elements of S less than 73.

In the present paper, we extend this table to include all the elements of S less than 1000. But our main purpose is first to give necessary conditions that k must satisfy in order to belong to S . Then we will describe a general method to find the squares which are sums of k consecutive squares, for any given k , this method being an application of the theory of Pell's equation. Furthermore, we will show that if k belongs to S , then there exist infinitely many squares that can be written as the sum of k consecutive squares if and only if k itself is not a square.

As usual, we will write $p^\alpha \parallel k$ iff $p^\alpha \mid k$ but $p^{\alpha+1} \nmid k$; α always will denote a strictly positive integer.

2. SOME NECESSARY CONDITIONS ON k

Lemma 0. *If $x^2 + y^2 \equiv 0 \pmod{2^{2n}}$, then $2^n \mid x$ and $2^n \mid y$.*

Proof: The assertion is trivial for $n = 1$. Suppose it is true for $n - 1$. $x^2 + y^2 \equiv 0 \pmod{2^{2n}}$ implies $x^2 + y^2 \equiv 0 \pmod{2^{2n-2}}$. By the induction hypothesis, $x = 2^{n-1}a$ and $y = 2^{n-1}b$, and so $2^{2n-2}a^2 + 2^{2n-2}b^2 \equiv 0 \pmod{2^{2n}}$. Dividing by 2^{2n-2} , we get $a^2 + b^2 \equiv 0 \pmod{4}$ from which we deduce that a and b are even. Thus $2^n \mid x$ and $2^n \mid y$.

Proposition. *Let k be an element of S .*

- (1) *If $2^\alpha \parallel k$, then α is odd.*
- (2) *If $3^\alpha \parallel k$, then α is odd.*
- (3) *If p is a prime, $p > 3$ and $p^\alpha \parallel k$ with α odd, then $p \equiv \pm 1 \pmod{12}$.*
- (4) *If p is a prime, $p > 3$ and $p^\alpha \parallel k + 1$ with $p \equiv 3 \pmod{4}$, then α is even.*
- (5) *If $3^\alpha \parallel k + 1$, then α is odd.*
- (6) *$k \not\equiv 3 \pmod{9}$.*
- (7) *$k \not\equiv 2^\alpha - 1$ or $2^\alpha \pmod{2^{\alpha+2}}$ for any $\alpha \geq 2$.*

Proof: Since $k \in S$, there exists $n \in \mathbb{N}$ such that $\sum_{i=1}^k (n+i)^2 = \square$. This can be rewritten in several different ways:

$$kn^2 + k(k+1)n + \frac{k(k+1)(2k+1)}{6} = \square \quad (i)$$

$$kn(n+k+1) + \frac{k(k+1)(2k+1)}{6} = \square \quad (ii)$$

$$9k(2n+k+1)^2 + 3k(k^2-1) = \square. \quad (iii)$$

To prove (1), suppose that $k = 2^\alpha r$ with α even ($\alpha \geq 2$) and r odd. Then, by (iii), $9 \cdot 2^\alpha r(2n + 2^\alpha r + 1)^2 + 3 \cdot 2^\alpha r(2^{2\alpha} r^2 - 1) = \square$ and since 2^α is a square, $9r(2n + 2^\alpha r + 1)^2 + 3r(2^{2\alpha} r^2 - 1) = \square$. Reducing modulo 4 we get $2 \equiv \square \pmod{4}$, a contradiction.

The proof of (2) is similar: suppose that $k = 3^\alpha r$ with α even ($\alpha \geq 2$) and $3 \nmid r$. Then, by (iii), $3^{\alpha+2} r(2n + 3^\alpha r + 1)^2 + 3^{\alpha+1} r(3^{2\alpha} r^2 - 1) = \square$ and since 3^α is a square, $9r(2n + 3^\alpha r + 1)^2 + 3r(3^{2\alpha} r^2 - 1) = \square$. Reducing modulo 9 we get $6r \equiv \square \pmod{9}$, a contradiction since $3 \nmid r$.

(3) Let p be a prime > 3 such that $p^\alpha \parallel k$ with α odd. Since α is odd and $k[9(2n+k+1)^2 + 3(k^2-1)] = \square$, p must divide the second factor on the left hand side, and so $9(2n+1)^2 \equiv 3 \pmod{p}$. Since $p \neq 3$, $3(2n+1)^2 \equiv 1 \pmod{p}$ and so the Legendre symbol $(3/p) = 1$, that is $p \equiv \pm 1 \pmod{12}$.

(4) Let p be a prime > 3 such that $p^\alpha \parallel k+1$ with $p \equiv 3 \pmod{4}$. Then $k+1 = ap^\alpha$ with $p \nmid a$, and so $(ap^\alpha - 1)[9(2n + ap^\alpha)^2 + 3ap^\alpha(ap^\alpha - 2)] = \square$ from which it follows that $-(6n)^2 \equiv \square \pmod{p}$. If $p \nmid n$, then $(-1/p) = 1$ (since $p \nmid 6$), contradicting $p \equiv 3 \pmod{4}$. Therefore $n = bp^\beta$ with $p \nmid b$ and $\beta > 0$ and we have

$$(ap^\alpha - 1)[9(2bp^\beta + ap^\alpha)^2 + 3ap^\alpha(ap^\alpha - 2)] = \square \quad (*)$$

If $\alpha < 2\beta$, then $(ap^\alpha - 1)p^\alpha[9(4b^2p^{2\beta-\alpha} + 4abp^\beta + a^2p^\alpha) + 3a(ap^\alpha - 2)] = \square$; in this product, the only factor divisible by p is p^α , and so α is even. If $\alpha = 2\beta$, then α is obviously even. Suppose now $\alpha > 2\beta$; dividing $(*)$ by $p^{2\beta}$, we get $(ap^\alpha - 1)[9(2b + ap^{\alpha-2\beta})^2 + 3ap^{\alpha-2\beta}(ap^\alpha - 2)] = \square$ and reducing modulo p , $-(6b)^2 \equiv \square \pmod{p}$. Since $p \nmid 6b$, we have $(-1/p) = 1$, contradicting again $p \equiv 3 \pmod{4}$, and so $\alpha > 2\beta$ is impossible.

(5) Let $k+1 = 3^\alpha a$ where $3 \nmid a$ and $\alpha \geq 2$. Then $(3^\alpha a - 1)[(2n + 3^\alpha a)^2 + 3^{\alpha-1}a(3^\alpha a - 2)] = \square$, from which we deduce that $-(2n)^2 \equiv \square \pmod{3}$, and so $n = 3^\beta b$, where $3 \nmid b$ and $\beta > 0$. Therefore

$$(3^\alpha a - 1)[(2 \cdot 3^\beta b + 3^\alpha a)^2 + 3^{\alpha-1}a(3^\alpha a - 2)] = \square.$$

Putting $\alpha' = \alpha - 1$, the same argument as in (4) (comparing α' and 2β) shows that α' is even, and so α is odd.

(6) Starting from equality (ii), suppose that $4kn(n+k+1) + (2k(k+1)(2k+1)/3) = \square$ where $k = 9\mu + 3$. Then, reducing modulo 3, we get $2(3\mu + 1)(9\mu + 4)(18\mu + 7) \equiv \square \pmod{3}$, that is $2 \equiv \square \pmod{3}$, a contradiction.

(7) Suppose that

$$9kn(n+k+1) + \frac{3k(k+1)(2k+1)}{2} = \square$$

where $k = 2^\alpha + \mu 2^{\alpha+2}$ ($\alpha \geq 2$). Reducing modulo $2^{\alpha+1}$, we get $3 \cdot 2^{\alpha-1} \equiv \square$

(mod $2^{\alpha+1}$). If α is odd, then $\alpha - 1$ is even and $3 \equiv \square \pmod{4}$. If α is even, then division by $2^{\alpha-2}$ yields $6 \equiv \square \pmod{8}$. In each case we have a contradiction. The second case in (7), i.e. $k = 2^\alpha - 1 + \mu 2^{\alpha+2}$ yields $3 \cdot 2^{\alpha-1} - 9n^2 \equiv \square \pmod{2^{\alpha+1}}$. Suppose then that $m^2 + 9n^2 \equiv 3 \cdot 2^{\alpha-1} \pmod{2^{\alpha+1}}$, which implies $m^2 + 9n^2 \equiv 0 \pmod{2^{\alpha-1}}$. If α is odd, Lemma 0 forces $m = 2^{(\alpha-1)/2}a$ and $n = 2^{(\alpha-1)/2}b$, and so $2^{\alpha-1}a^2 + 9 \cdot 2^{\alpha-1}b^2 \equiv 3 \cdot 2^{\alpha-1} \pmod{2^{\alpha+1}}$, from which we deduce that $a^2 + b^2 \equiv 3 \pmod{4}$, a contradiction. If $\alpha = 2$, then $m^2 + n^2 \equiv 6 \pmod{8}$, a contradiction. If α is even and $\alpha > 2$, we have also $m^2 + 9n^2 \equiv 0 \pmod{2^{\alpha-2}}$, and so $m = 2^{(\alpha-2)/2}a$ and $n = 2^{(\alpha-2)/2}b$, which gives $2^{\alpha-2}a^2 + 9 \cdot 2^{\alpha-2}b^2 \equiv 3 \cdot 2^{\alpha-1} \pmod{2^{\alpha+1}}$ and $a^2 + b^2 \equiv 6 \pmod{8}$, another contradiction.

3. THE ELEMENTS OF S LESS THAN 1000. The following table gives the values of k less than 1000 for which there exists a sum of k consecutive squares equal to a square and, for each of these values of k , the smallest value of $n + 1$ for which $(n + 1)^2 + \cdots + (n + k)^2 = \square$.

k	$n + 1$	k	$n + 1$	k	$n + 1$	k	$n + 1$
1	1	184	7	383	9985	625	301
2	3	191	4493	393	802	649	38
11	18	193	83342	407	183	673	177
23	7	194	83	409	71752	674	126031
24	1	218	65	431	10123	698	322
26	25	239	899	443	16806	722	131
33	7	241	3807	457	94707486	753	197
47	539	242	64	458	1081	767	6665193
49	25	249	556	479	7989	793	516232
50	7	289	20	481	1663	794	1122
59	22	297	106	491	11476	841	1678
73	442	299	132	506	1778	856	8617
74	225	311	2277	529	255	863	23172625
88	192	312	15	537	68680	864	65
96	13	313	1788	539	172	866	12116
97	15	337	5063	554	25	887	413
107	26914	338	27	568	443	897	454
121	244	347	11320	577	167065	913	688013
122	50	352	280	578	3	914	3480
146	5552	361	358	587	310097726	961	3
169	30	362	1805	599	2927	971	51958
177	553	376	210	600	25	983	9977

4. SOME COMMENTS. The 7 necessary conditions given in the proposition exclude immediately 910 values of $k \leq 1000$. The 90 remaining values are the 88 ones given in the preceding table (for which there is a solution of the equation $(n + 1)^2 + \cdots + (n + k)^2 = \square$), as well as two sporadic values $k = 25$ and $k = 842$ for which there is no solution, but which nevertheless satisfy the 7 conditions. Thus for $k \leq 1000$, $k \neq 25$ and $k \neq 842$, the elements of S and the values of k satisfying the 7 conditions coincide. Are there other simple necessary conditions to be added to those seven in order to obtain a set of necessary and sufficient conditions? I could not find quick arguments showing that 25 and $842 \notin S$, but this will be proved in the next section as an application of a method of resolution of Pell's equation $x^2 - ky^2 = a$ (see examples 2 and 4 below), the equation $(n + 1)^2 + \cdots + (n + k)^2 = \square$ being a special case.

5. PELL'S EQUATION. Let

$$x^2 - ky^2 = a \quad (P)$$

where $k \in \mathbb{N}$ and $a \in \mathbb{Z}$. We will consider only the solutions of (P) for which x and y are positive integers. Consider first the case where k is a square. Then $(x - \sqrt{ky})(x + \sqrt{ky}) = a$, which has only a finite number of solutions: each pair u, v of integers such that $uv = a$, $u \geq v$ and $2\sqrt{k} \mid (u - v)$ yields a solution $x = (u + v)/2$ and $y = (u - v)/2\sqrt{k}$.

Example 1. To find the squares which can be expressed as a sum of 49 consecutive squares, we must solve the Diophantine equation $49n^2 + 2450n + 40425 = m^2$ (see equality (i)). Putting $m = 7x$, we get, $n^2 + 50n + 825 = x^2$, which we reduce to Pell's equation $x^2 - y^2 = 200$ by putting $y = n + 25$. Its solutions are $(x, y) = (15, 5), (27, 23), (51, 49)$. Since n must be positive, we have $y \geq 25$ and so the only solution is $n = 24$ and $m = 357$; this gives the only sum of 49 consecutive squares which is a square, namely $25^2 + \cdots + 73^2 = 357^2$.

Example 2. To show that there is no square which is a sum of 25 consecutive squares, we write (as in Example 1) the equation $25n^2 + 650n + 5525 = m^2$, and we reduce it to $x^2 - y^2 = 52$ by putting $m = 5x$ and $y = n + 13$. Then it is easy to check that the latter equation has no solution with $n \geq 0$.

Suppose now that k is not a square. If (x_1, y_1) and (x_2, y_2) are two different solutions of (P), we will write $(x_1, y_1) < (x_2, y_2)$ if $x_1 < x_2$ and $y_1 < y_2$.

Lemma 1. *The solutions of (P) can be ordered.*

Proof: Let (x_1, y_1) and (x_2, y_2) be two different solutions of (P). Then $x_1^2 - ky_1^2 = x_2^2 - ky_2^2 = a$, which implies $x_1^2 - x_2^2 = k(y_1^2 - y_2^2)$, and so $x_1 - x_2$ and $y_1 - y_2$ are both positive or both negative. So either $(x_1, y_1) < (x_2, y_2)$ or $(x_2, y_2) < (x_1, y_1)$.

Lemma 2. *The equation $x^2 - ky^2 = 1$ has always a solution different from $(1, 0)$.*

For a proof of this well known result, we refer the reader to [3].

Lemma 3. *Let (λ, μ) be the smallest solution of $x^2 - ky^2 = 1$ different from $(1, 0)$ and let M denote the matrix*

$$\begin{bmatrix} \lambda & k\mu \\ \mu & \lambda \end{bmatrix}.$$

Then, if (x, y) is a solution of (P), $M^k(x, y)$ is also a solution for every $k \in \mathbb{Z}$, where M^0 is the 2×2 unit matrix. (Note: for notational convenience, we sometimes write row vectors instead of column vectors.)

Proof: Since $M(x, y) = (\lambda x + k\mu y, \mu x + \lambda y)$ and

$$M^{-1} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \lambda & -k\mu \\ -\mu & \lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \lambda x - k\mu y \\ -\mu x + \lambda y \end{bmatrix}$$

we only have to show that $(\lambda x + \varepsilon k\mu y, \varepsilon \mu x + \lambda y)$ with $\varepsilon = \pm 1$ is a solution of (P), which is quite easy.

The proofs of the following two lemmas are straightforward:

Lemma 4. *If $(x_1, y_1) < (x_2, y_2)$, then $M(x_1, y_1) < M(x_2, y_2)$ and $M^{-1}(x_1, y_1) < M^{-1}(x_2, y_2)$.*

Lemma 5. $M^{-1}(x, y) < (x, y) < M(x, y)$.

CONCLUSION. The preceding results suggest the following method to solve (P). First, if $a > 0$, the smallest (real) solution of (P) is $(\sqrt{a}, 0)$. Since $M(\sqrt{a}, 0) = \sqrt{a}(\lambda, \mu)$, if we can find all the integer solutions of (P) lying between $(\sqrt{a}, 0)$ and $\sqrt{a}(\lambda, \mu)$, say $(x_1, y_1), \dots, (x_r, y_r)$, then the set of all solutions of (P) is given by $\{M^\alpha(x_i, y_i) | \alpha \in N, i = 1, \dots, r\}$. Likewise, if $a < 0$, it suffices to determine all the solutions (x_i, y_i) between

$$(0, \sqrt{-\frac{a}{k}}) \quad \text{and} \quad M(0, \sqrt{-\frac{a}{k}}) = \sqrt{-\frac{a}{k}}(k\mu, \lambda).$$

In order to apply this method, we have to determine the smallest nontrivial solution (λ, μ) of $x^2 - ky^2 = 1$, which can be found by using continued fractions ([4]). The table of Section 6 below shows that the distribution of (λ, μ) seems to be rather chaotic.

Example 3. To find the squares which can be expressed as a sum of 11 consecutive squares, we have to solve the equation $11n^2 + 132n + 506 = m^2$. Putting $m = 11y$ and $x = n + 6$, we get $x^2 - 11y^2 = -10$. For $k = 11$, we have $\lambda = 10$ and $\mu = 3$, and so it suffices to find all the solutions between

$$\sqrt{\frac{10}{11}}(0, 1) \quad \text{and} \quad \sqrt{\frac{10}{11}}(33, 10)$$

i.e. the solutions (x, y) with $1 \leq y \leq 9$; these are $(1, 1)$ and $(23, 7)$. All the other solutions are given by

$$\begin{aligned} \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 10 & 33 \\ 3 & 10 \end{bmatrix}^\alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \frac{1}{2} \left[\begin{array}{c} (10 + 3\sqrt{11})^\alpha (1 + \sqrt{11}) + (10 - 3\sqrt{11})^\alpha (1 - \sqrt{11}) \\ \frac{1}{\sqrt{11}} (10 + 3\sqrt{11})^\alpha (1 + \sqrt{11}) - \frac{1}{\sqrt{11}} (10 - 3\sqrt{11})^\alpha (1 - \sqrt{11}) \end{array} \right] \end{aligned}$$

and

$$\begin{aligned} \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 10 & 33 \\ 3 & 10 \end{bmatrix}^\alpha \begin{bmatrix} 23 \\ 7 \end{bmatrix} \\ &= \frac{1}{2} \left[\begin{array}{c} (10 + 3\sqrt{11})^\alpha (23 + 7\sqrt{11}) + (10 - 3\sqrt{11})^\alpha (23 - 7\sqrt{11}) \\ \frac{1}{\sqrt{11}} (10 + 3\sqrt{11})^\alpha (23 + 7\sqrt{11}) - \frac{1}{\sqrt{11}} (10 - 3\sqrt{11})^\alpha (23 - 7\sqrt{11}) \end{array} \right]. \end{aligned}$$

Example 4. To show that there is no square which is a sum of 842 consecutive squares, we must solve $842n^2 + 709806n + 199337185 = m^2$. Putting $x = 2n + 843$, $m = 421y$, we get $x^2 - 842y^2 = -236321$. For $k = 842$, we have $\lambda = 1683$ and $\mu = 58$; so it suffices to find the solutions (x, y) with $0 \leq x \leq 818154$ and $17 \leq y \leq 28195$, but an easy computer check shows that no solution exists in this interval (this check can be made very fast by noticing that $x^2 \equiv 281 \pmod{842}$, and

so $x \equiv \pm 165 \pmod{842}$, from which it follows that only the integers x of the form $842t \pm 165$ with $0 \leq t \leq 971$ must be considered).

6. THE SMALLEST NON TRIVIAL SOLUTION(λ, μ) OF $x^2 - ky^2 = 1$

k	λ	μ	k	λ	μ	k	λ	μ
2	3	2	30	11	2	56	15	2
3	2	1	31	1520	273	57	151	20
5	9	4	32	17	3	58	19603	2574
6	5	2	33	23	4	59	530	69
7	8	3	34	35	6	60	31	4
8	3	1	35	6	1	61	1766319049	226153980
10	19	6	37	73	12	62	63	8
11	10	3	38	37	6	63	8	1
12	7	2	39	25	4	65	129	16
13	649	180	40	19	3	66	65	8
14	15	4	41	2049	320	67	48842	5967
15	4	1	42	13	2	68	33	4
17	33	8	43	3482	531	69	7775	936
18	17	4	44	199	30	70	251	30
19	170	39	45	161	24	71	3480	413
20	9	2	46	24335	3588	72	17	2
21	55	12	47	48	7	73	2281249	267000
22	197	42	48	7	1	74	3699	430
23	24	5	50	99	14	75	26	3
24	5	1	51	50	7	76	57799	6630
26	51	10	52	649	90	77	351	40
27	26	5	53	66249	9100	78	53	6
28	127	24	54	485	66	79	80	9
29	9801	1820	55	89	12	80	9	1

REFERENCES

1. W. S. Anglin, The Square Pyramid Puzzle, *Amer. Math. Monthly* 97 (1990), 120–124.
2. E. Lucas, Question 1180, *Nouvelles Annales de Mathématiques*, ser. 2, 14 (1875), 336.
3. L. Mordell, *Diophantine Equations*, Academic Press, London (1969).
4. I. Niven, H. S. Zuckerman, H. L. Montgomery, *An Introduction to the Theory of Numbers*, Wiley, New-York (1991).
5. G. N. Watson, The Problem of the Square Pyramid, *Messenger of Mathematics* 48 (1918–19), 1–22.
6. Squares Expressible as a Sum of n Consecutive Squares, Advanced Problem 6552, proposed by M. Laub, *Amer. Math. Monthly* 97 (1990), 622–625.

*Département de Mathématiques
Université Libre de Bruxelles
Campus Plaine C. P. 216
B-1050 Bruxelles, Belgium*

Square Roots mod p

Stephen M. Turner

We would like, given an odd prime p , and an integer y , to solve

$$x^2 \equiv y \pmod{p}. \quad (1)$$

There are fast ways of determining whether there *is* a solution—if p does not divide y , *Euler's criterion* says that there is a solution precisely when

$$y^{(p-1)/2} \equiv 1 \pmod{p} \quad (2)$$

and the calculation can be made more manageable using the so-called *law of quadratic reciprocity* (see almost any book on elementary number theory—like Rose's [1], in the list at the end). If p does divide y , then we have a unique solution $x \equiv 0 \pmod{p}$.

Suppose then, that we have the equation (1) as above, where p is an odd prime that does not divide y , and that there is known to be a solution. How do we go about finding one? Finding one enables the equation to be solved completely, as we can find the only other solution by changing sign (mod p). Let us call solutions of (1) *square roots* (mod p). Values of $y \pmod{p}$ for which (1) has solutions are usually called *quadratic residues* (mod p).

From now on, all congruences are taken to be (mod p). In some cases, solutions are easy to find: if we can find a positive odd integer C such that

$$y^C \equiv 1$$

then we can just take either of

$$x \equiv \pm y^{(C+1)/2} \quad (3)$$

for then we have

$$x^2 \equiv y^{C+1} \equiv y$$

and we are done. This is always possible if $p \equiv 3 \pmod{4}$, for then we can take $C = (p-1)/2$, using Euler's criterion again. This choice for C is no good if $p \equiv 1 \pmod{4}$, but at least we have got a solution which works half of the time. We may be able to find a different C , though—see later.

Note that we can only find such a C if there is a solution of the form $x \equiv (\text{power of } y)$, which need not be the case in general. (Examples in a moment). This point is made by Gauss in his investigation of this problem (see [5]), and he goes on to give a procedure which sometimes simplifies it in other cases.

The first general method of solution appears to be due to Shanks (1972). This method (see [2]) starts with a simple rule for choosing an odd number C , just as above, which either gives a solution immediately or can be used to generate one.

Shanks suggests choosing C to be the largest odd divisor of $p-1$: i.e., choose C by writing

$$p-1 = C \cdot 2^s \quad (4)$$

where C is odd and s is positive. When $p \equiv 3 \pmod{4}$, this gives the same value for C as before, so we have generalized Gauss's construction.

Example. Take $p = 13$, so that $C = 3$. As $13 \equiv 1 \pmod{4}$, Gauss's method is inapplicable. Euler's criterion shows that both 3 and -1 have square roots mod 13. Trying the above method.

(i) $y \equiv 3$: then $y^C \equiv 3^3 \equiv 1 \pmod{13}$. Take $x \equiv y^2 \equiv 9$. And we have found the solutions $x \equiv \pm 9$.

(ii) $y \equiv -1$: then $y^C \equiv (-1)^3 \equiv -1 \pmod{13}$ and the method fails, although there are solutions $x \equiv \pm 5$.

The second part of the example shows that it is not always possible to find an odd C with the required property, since we always have square roots of -1 when $p \equiv 1 \pmod{4}$ but $(-1)^C \equiv -1 \pmod{p}$ for any odd C (and $-1 \not\equiv 1$ as p is odd!).

As mentioned already, Shanks uses this C in the first step of an algorithm which solves (1) completely in all cases. In his article, he claims that his choice for C solves the congruence immediately in $2/3$ of all cases. This note makes this claim precise, and proves it. The proof requires very little previous knowledge: any results needed are quoted, and they seem quite interesting in their own right anyway.

We certainly expect that the method solves (1) immediately in at least $\frac{1}{2}$ of all cases, as half of the primes are congruent to $3 \pmod{4}$.

Let p and y be as above, and let C and s be defined as in (4). Let E denote the event 'congruence solved immediately' and write 'Prob' for probability. The proof of Shanks's claim goes like this:

- (a) we find $\text{Prob}(E, \text{ for some fixed odd } p, \text{ but randomly chosen } y)$;
- (b) we choose a number $M > 0$, and find an expression for $\text{Prob}(E, \text{ given odd } p \leq M, \text{ but randomly chosen } y)$ —given that y has square roots \pmod{p} ;
- (c) we let M approach infinity, and show that we can define what we mean by $\text{Prob}(E)$ in an unambiguous way—where p and y are chosen randomly subject to the usual conditions—and that the limiting value is $\frac{2}{3}$, establishing Shanks's claim.

PART (A). In this section we use the following theorem about cyclic groups (a *cyclic* group is a group consisting entirely of the powers of one of its elements, the element being a *generator* of the group.)

Theorem 1. *Let G be a cyclic group with m elements, and suppose that n divides m . Then there are exactly n elements in G satisfying $x^n = \text{identity of } G$.*

Given a fixed odd prime p , we know that there are $p - 1$ residue classes which are not divisible by p : in fact these form a cyclic group G (say). Since each quadratic residue has exactly two distinct square roots, there are exactly $(p - 1)/2$ quadratic residues \pmod{p} , and it is not hard to see that these form a subgroup H (say) of G . Then the probability we require is simply given by (say)

$$f(p) = \frac{\#\{y \text{ in } H: y^C = \text{identity}\}}{\#\{y \text{ in } H\}}$$

Noting that H is itself a cyclic group of $(p - 1)/2$ elements, we apply Theorem 1 (noting that C divides $\#H$) to get

$$f(p) = \frac{C}{(p - 1)/2} = \frac{2}{2^s}. \quad (5)$$

Then *define* the function f on integers m ($m \neq 1$) by

$$f(m) = 2A/(m-1) \quad (5a)$$

where A is the largest odd divisor of $m-1$, so that we have generalized f ; this will be useful later on. This concludes part (a).

PART (B). Fix some $M \geq 3$, and write $P(M) = \{\text{odd primes} \leq M\}$. If we choose p randomly from $P(M)$, and then a random quadratic residue $y \pmod{p}$, we have (say)

$$T(M) = \text{Prob}(E \text{ given } p \text{ in } P(M)) = \sum_{q \in P(M)} \text{Prob}(E \text{ and } p = q) \quad (6)$$

where the probabilities add as the events are exclusive, and using the definition of conditional probability, namely that for any two events A and B ,

$$\text{Prob}(A \text{ given } B) = \text{Prob}(A \text{ and } B) / \text{Prob}(B),$$

we can write

$$T(M) = \sum_{q \in P(M)} \text{Prob}(E \text{ given } p = q) \text{Prob}(p = q)$$

and $\text{Prob}(E \text{ given } p = q)$ is just $f(q)$ in the notation of part (a), so we have

$$T(M) = \sum_{q \in P(M)} f(q) \text{Prob}(p = q). \quad (7)$$

For any q in $P(M)$, we have $f(q) = 2/(2^t)$ for some integer $t \geq 1$; now we can rewrite the sum in (7) by summing not over the elements in $P(M)$, but over the values attained by the function f —i.e. by grouping together m and n (say) precisely when $f(m) = f(n)$. So we get

$$T(M) = \sum_{t=1}^{\infty} \frac{2}{2^t} \text{Prob}\left(p \text{ in } P(M) \text{ and } f(p) = \frac{2}{2^t}\right) \quad (8)$$

which is a finite series, as $2^t > M$ for large enough t . We want now to find what proportion of the odd primes q in $P(M)$ satisfy $f(q) = 2/(2^t)$ for any given t . Using the Definition (5a) of the function f , we see that for any integer m (except 1), and for any integer $t \geq 0$,

$$f(m) = 2/(2^t) \text{ precisely when } m \equiv (1 + 2^t) \pmod{2^{t+1}}, \quad (9)$$

since we must be able to write $m-1 = A \cdot 2^t$ when this is so, for some *odd* A . In general there is no way of calculating the proportion of *primes* in $P(M)$, which have this property except by counting them, so we need something a bit more sophisticated if we are to make any progress—that is if we hope to be able to write down an expression for such a quantity. We can however find an approximation which is good for sufficiently large M , which is all that we need anyway. This way forward is used in the next section.

PART (C). We need the following well-known result, which is Dirichlet's theorem on primes in arithmetic progression.

Theorem 2. *Let a and b be coprime, with $b > 0$. Then there are infinitely many primes congruent to $a \pmod{b}$ (i.e. the arithmetic progression $a, a+b, a+2b, \dots$ contains infinitely many primes).*

In fact, a stronger version of Theorem 2 will be useful, to state which we use the following definition. *Euler's phi-function* is defined as follows: if $n > 0$, $\phi(n)$ is defined to be the number of congruence classes coprime to n —so that $\phi(10)$ is 4, since $\{1, 3, 7, 9\}$ are coprime to 10, but $\{0, 2, 4, 5, 6, 8\}$ are not. We will use the facts that

(i) $\phi(2^{t+1}) = 2^t$; and

(ii) in Theorem 2, given b , there are $\phi(b)$ choices for a giving different progressions.

The result we need strengthens Theorem 2 by stating that the primes which are coprime to b are distributed roughly equally among the progressions.

Theorem 3. *With the same a and b , let R be positive. Then*

$$\frac{\#\{p: p \text{ prime}, p \leq R, p \equiv a \pmod{b}\}}{\#\{p: p \text{ prime}, p \leq R\}} \rightarrow \frac{1}{\phi(b)}$$

as R goes to infinity. (See [3] for a proof).

We can think of the quantity on the left in Theorem 3 as the probability that a randomly chosen prime which does not exceed R is congruent to $a \pmod{b}$. In fact, we can suppose that we exclude from the denominator any primes which divide b —we will want to do this in practice—and the limiting value on the right in Theorem 3 is unaffected, since b has only finitely many prime divisors. Such primes are already excluded from the numerator, as a and b are coprime. We will write

$$G(a, b, R) = \frac{\#\{p: p \text{ prime}, p \leq R, p \equiv a \pmod{b}\}}{\#\{p: p \text{ prime}, p \leq R, p \text{ coprime to } b\}} \quad (10)$$

Putting $b = 2^{t+1}$, $a = 1 + 2^t$, $R = M$, we get, using (9)

$$G(1 + 2^t, 2^{t+1}, M) = \frac{\#\{p: p \text{ in } P(M), p \equiv 1 + 2^t \pmod{2^{t+1}}\}}{\#\{p: p \text{ in } P(M)\}}$$

as only the prime 2 needs to be excluded, and this is just the same as

$$\text{Prob}(p \text{ in } P(M) \text{ and } f(p) = 2/(2^t))$$

the quantity which appears in (8). So we can write (8) as

$$T(M) = \sum_{t=1}^{\infty} \frac{2}{2^t} G(1 + 2^t, 2^{t+1}, M), \quad (11)$$

recalling that $T(M) = \text{Prob}(E \text{ given } p \text{ in } P(M))$. What we would like to do is to evaluate $\lim_{M \rightarrow \infty} T(M)$, and call this $\text{Prob}(E)$. There is a snag however, because we have to interchange the two limiting processes of “ $M \rightarrow \infty$ ” and the infinite summation in (11) to perform the calculation, as we cannot find $T(M)$ directly. To express things more succinctly, we can write (11) itself as a limit, i.e. write $T(M) = \lim_{N \rightarrow \infty} A(M, N)$, where we define

$$A(M, N) = \sum_{t=1}^N \frac{2}{2^t} G(1 + 2^t, 2^{t+1}, M) \quad (12)$$

$A(M, N)$ is an example of a double sequence—that is, it is indexed by two sets like $\{1, 2, 3, \dots\}$. Note that since the primes themselves form a *discrete* set in the reals (a set consisting of isolated points), we can suppose that M is an integer, without affecting the limiting behaviour as $M \rightarrow \infty$.

Thus we want to define $\text{Prob}(E) = \lim_{M \rightarrow \infty} (\lim_{N \rightarrow \infty} A(M, N))$, but cannot evaluate this directly. We can, however, easily find $\lim_{N \rightarrow \infty} (\lim_{M \rightarrow \infty} A(M, N))$, the *iterated limit* in the other order since we have

$$\lim_{M \rightarrow \infty} A(M, N) = \sum_{t=1}^N \frac{2}{2^t} \left(\frac{1}{\text{phi}(2^{t+1})} \right) = \frac{2}{3} \left(1 - \frac{1}{4^N} \right)$$

and so

$$\lim_{N \rightarrow \infty} \left(\lim_{M \rightarrow \infty} A(M, N) \right) = 2/3. \quad (13)$$

Unfortunately, we cannot interchange the order of iteration without justification, since, for example

$$\lim_{M \rightarrow \infty} \left(\lim_{N \rightarrow \infty} \frac{M}{M+N} \right) = 0$$

but

$$\lim_{N \rightarrow \infty} \left(\lim_{M \rightarrow \infty} \frac{M}{M+N} \right) = 1.$$

We will see that in our case, the interchange can be justified. Firstly, we define, for a double sequence $B(M, N)$, the *double limit* (if it exists) to be

$$\lim_{K \rightarrow \infty} B(M, N), \quad \text{where } K = \text{minimum}(M, N). \quad (14)$$

Note that a double limit may exist even if neither of the iterated limits does: to see this, consider the example

$$B(M, N) = \frac{(-1)^M}{N} + \frac{(-1)^N}{M}.$$

In our case, we will have, for the sequence $A(M, N)$ defined above, that the double limit and the two iterated limits all exist and agree, with the common value $\frac{2}{3}$, so proving Shanks's claim. We will need the following (see [4]).

Theorem 4. *Let $B(M, N)$ be any double sequence of complex numbers. Suppose that*

- (i) $\lim_{M \rightarrow \infty} B(M, N)$ exists for all N ;
- (ii) $\lim_{N \rightarrow \infty} B(M, N)$ exists for all M ; and
- (iii) *the double limit exists.*

Then the two iterated limits exist and agree.

From this, it follows quite readily that the iterated limits also agree with the double limit. In our case, this means that $\text{Prob}(E)$ exists and is defined unambiguously, with the value $\frac{2}{3}$ (since this is the value found in (13)). Hence, Shanks's claim will follow once we have shown that $A(M, N)$ satisfies the hypotheses of Theorem 4. Checking this:

- (i) already shown (in deriving (13));
- (ii) for each fixed M , $A(M, N)$ is just a finite sum as $G(1 + 2^t, 2^{t+1}, M) = 0$ for all large enough t (see (8));
- (iii) we will actually show directly that the double limit is $\frac{2}{3}$. This is done as follows:

It is enough to show that by choosing both M and N to be sufficiently large, we can make $A(M, N) - \frac{2}{3}$ arbitrarily small in absolute value. Fix an $\varepsilon > 0$, and

choose N_0 so that

$$N \geq N_0 \Rightarrow 2^N \geq 6/\varepsilon \quad (15)$$

where this comes from the two conditions which we will (with the benefit of hindsight!) require, namely: $1 - 1/4^N \geq 1 - \frac{1}{2}\varepsilon$, whenever $N \geq N_0$, and

$$\sum_{t=N_0+1}^{\infty} \left(\frac{2}{2^t} \right) \leq \frac{1}{3}\varepsilon.$$

We know that, for each t , $\lim_{M \rightarrow \infty} G(1 + 2^t, 2^{t+1}, M) = 2^{-t}$ so for each t in $\{1, 2, 3, \dots, N_0\}$, there exists an M_t such that

$$M \geq M_t \Rightarrow \frac{1 - \varepsilon}{2^t} \leq G(1 + 2^t, 2^{t+1}, M) \leq \frac{1 + \varepsilon}{2^t}. \quad (16)$$

Then write $M_0 = \max(M_1, M_2, \dots, M_{N_0})$, so that, whenever $M \geq M_0$, substituting (16) into (12), we get (after summing the geometric series which results), that

$$(1 - \varepsilon) \frac{2}{3} \left(1 - \frac{1}{4^{N_0}} \right) \leq A(M, N_0) \leq (1 + \varepsilon) \frac{2}{3} \left(1 - \frac{1}{4^{N_0}} \right). \quad (17)$$

By construction of N_0 , we have

$$1 - \frac{1}{2}\varepsilon \leq 1 - \frac{1}{4^{N_0}} \leq 1,$$

and substituting this into (17) gives (since $\frac{1}{3}\varepsilon^2 > 0$)

$$-\varepsilon + \frac{2}{3} \leq A(M, N_0) \leq \frac{2}{3} + \frac{2}{3}\varepsilon. \quad (18)$$

Now suppose that $N \geq N_0$, and write $A(M, N) = A(M, N_0) + Z$, where

$$Z = \sum_{t=N_0+1}^N \frac{2}{2^t} G(1 + 2^t, 2^{t+1}, M). \quad (19)$$

Since the G 's are probabilities, and by our choice of N_0 ,

$$0 \leq Z \leq \sum_{t=N_0+1}^N \frac{2}{2^t} \leq \frac{1}{3}\varepsilon.$$

Finally, then if $M \geq M_0$, and $N \geq N_0$, and so certainly whenever $\min(M, N) \geq \max(M_0, N_0)$,

$$-\varepsilon + \frac{2}{3} \leq Z - \varepsilon + \frac{2}{3} \leq A(M, N) \leq Z + \frac{2}{3} + \frac{2}{3}\varepsilon \leq \varepsilon + \frac{2}{3}$$

and hence the double limit is $\frac{2}{3}$. This checks condition (iii) of Theorem 4 for the sequence $A(M, N)$ and so establishes Shanks's claim.

SUMMARY. We have proved the following: given an odd prime p , chosen at random, and a random quadratic residue $y \pmod{p}$, the square roots of $y \pmod{p}$ are, in $\frac{2}{3}$ of all cases, the values of x given by Equation (3), where C is the largest odd divisor of $p - 1$. (It is also true that we can generate the square roots from these starting values in other cases, as Shanks shows.)

REFERENCES

1. H. E. Rose, *A Course in Number Theory*, Oxford University Press, 1988, quadratic residues, pp. 51 foll.

2. D. Shanks, Five Number-theoretic Algorithms, *Proc. of 2nd Manitoba Conference on Numerical Mathematics*, edited by R. S. D. Thomas, and H. C. Williams, Utilitas Mathematical (Winnipeg). Shanks's name for the algorithm described above is RESSOL.
3. R. Ayoub, *An Introduction to the Analytic Theory of Numbers*, American Mathematical Society (1963). Our Theorem 3 is his Theorem 5.1 of Chapter 1.
4. J. C. Burkill and H. Burkill, *A Second Course in Mathematical Analysis*, Cambridge University Press, 1970. Our Theorem 4 is their Theorem 4.72 (i.e., in Chapter 4).
5. C. F. Gauss, *Disquisitiones Arithmeticae*, article 66, then articles 67-68. Published in English by Yale University Press, 1966, translated by A. A. Clarke.

Department of Mathematics
University of Glasgow
Glasgow G12 8QW
United Kingdom
smt@maths.gla.ac.uk

PICTURE PUZZLE
(from the collection of Paul Halmos)



A son more famous than his father?
 (see page 458.)

Estimate (6) shows that $b_N \leq 2\pi^2/3$. A simple calculation gives

$$b_N \geq \frac{4N}{2N+1} \left[\sum_{k=1}^{N+1} \frac{1}{k^2} \right] + \frac{4(N-1)}{2N+1} \\ \times \left[\frac{\frac{1}{1^2} + \left(\frac{1}{1^2} + \frac{1}{2^2} \right) + \cdots + \left(\frac{1}{1^2} + \frac{1}{2^2} + \cdots + \frac{1}{(N-1)^2} \right)}{N-1} \right].$$

Applying the squeeze law, we obtain that b_N tends to $2\pi^2/3$ as $N \rightarrow \infty$. Using (3) and (5), we obtain that for this choice of (a_n) , the ratio of the left-hand side of (1) and $(\sum_n a_n^2)^{1/2}$ converges to π as $N \rightarrow \infty$. This proves that π is the best possible constant in (1), Q.E.D.

Inequality (1) is known to be strict if (a_n) is nonzero. To see this for compactly supported nonzero sequences, observe that (6) is a strict bound for (5) since for some m and n , $2a_n a_m < a_n^2 + a_m^2$. For general sequences, a further argument is needed since the passage to the limit will destroy the strict inequality. See [HLP] for details on this.

To the best of my knowledge, the determination of the best constant for the l^p inequality, $1 < p \neq 2 < \infty$ remains unresolved. Pichorides [P] solves this problem for the corresponding continuous operator.

We end this note by asking a question: Is there a constant C such that for all square summable sequences (a_n) and all bounded sequences (λ_n) , the following inequality holds?

$$\left(\sum_{j \in \mathbf{Z}} \left| \sum_{\substack{n \in \mathbf{Z} \\ n \neq j}} \lambda_{j+n} \frac{a_n}{j-n} \right|^2 \right)^{1/2} \leq C \sup_j |\lambda_j| \left(\sum_{n \in \mathbf{Z}} |a_n|^2 \right)^{1/2}. \quad (7)$$

If $\lambda_n = 1$ for all n , then (7) reduces to (1) and one can take $C \geq \pi$.

REFERENCES

- [HLP] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge University Press (1934).
- [P] S. Pichorides, *On the Best Values of the Constants in the Theorems of M. Riesz, Zygmund and Kolmogorov*, *Studia Mathematica* 46 (1972), 164–179.
- [S] I. Schur, *Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen*, *Journal f. Math.* 140 (1911), 1–28.
- [W] H. Weyl, *Singuläre Integralgleichungen mit besonderer Berücksichtigung des Fourierschen Integraltheorems*, Doctoral Dissertation, University of Göttingen (1908).

Department of Mathematics
Washington University, Campus Box 1146
One Brookings Drive, St Louis, MO 63130-4899
grafakos@math.wustl.edu

Answer to Picture Puzzle
(p. 449)
Henri Cartan, the son of Eli Cartan.

NOTES

Edited by: John Duncan

Kummer's Test Gives Characterizations for Convergence or Divergence of all Positive Series

Jingcheng Tong

One of the basic facts about a positive series is that the series converges if and only if its partial sums are bounded. Equivalently, a positive series diverges if and only if its partial sums are unbounded. These observations, however, do not provide a practical way of determining convergence or divergence of positive series. Thus, it is surprising that Kummer's test, which is a source of many useful convergence and divergence tests, is equivalent to the boundedness of the partial sums, and it is also surprising that this fact is not well publicized and possibly unknown until now. Kummer's test is the sufficient part of our main result, which will be stated shortly.

There is a short introduction on the history of Kummer's test in [9]. The test was given by the German mathematician Ernst E. Kummer as early as 1835 (*Journ. reine angew. Math.* Vol 13, p. 172), though with a restrictive condition which was first recognized by U. Dini in 1867. Later it was rediscovered several times and gave rise as late as 1888, to a violent contention on questions of priority.

Kummer's test gives very powerful sufficient conditions for convergence or divergence of a positive series. As mentioned above it is the source of many other tests ([6], [10]). For instance, D'Alembert's test, Raabe's test, Bertrand's test, and Gauss' test are all special cases of Kummer's test obtained by choosing specific "parameters" p_k 's.

Theorem. Let $\sum a_k$ be a positive series.

- (1) $\sum a_k$ is convergent if and only if there is a positive series $\sum p_k$ and a real number $c > 0$, such that $p_k(a_k/a_{k+1}) - p_{k+1} \geq c$.
- (2) $\sum a_k$ is divergent if and only if there is a positive series $\sum p_k$, such that $\sum 1/p_k$ diverges and $p_k(a_k/a_{k+1}) - p_{k+1} \leq 0$.

Proof: Although proofs of the sufficiencies of (1), (2) can be found in many books on advanced calculus, we feel their simplicity warrants their inclusion for completeness.

Sufficiency.

(1) Suppose that $\sum p_k$ is a positive series and there is a real number $c > 0$, such that $p_k(a_k/a_{k+1}) - p_{k+1} \geq c$. This implies that

$$\begin{aligned} a_1 + \sum_{k=2}^N a_k &\leq a_1 + (1/c) \sum_{k=1}^N (p_k a_k - p_{k+1} a_{k+1}) \\ &= a_1 + (1/c) \left(\sum_{k=1}^N p_k a_k - \sum_{k=2}^{N+1} p_k a_k \right) \\ &= a_1(1 + p_1/c) - \frac{p_{N+1} a_{N+1}}{c} \leq a_1(1 + p_1/c). \end{aligned}$$

(2) Suppose that $\sum p_k$ is a positive series for which $\sum 1/p_k$ diverges and $p_k(a_k/a_{k+1}) - p_{k+1} \leq 0$. This implies that

$$a_2 \geq \frac{p_1 a_1}{p_2}, a_3 \geq \frac{p_2 a_2}{p_3} \geq \frac{p_1 a_1}{p_3}, \dots,$$

and consequently,

$$\sum_{k=1}^{\infty} a_k \geq a_1 + (p_1 a_1) \sum_{k=2}^{\infty} (1/p_k).$$

Necessity.

(1) Suppose that $\sum a_k$ is a convergent positive series. We are going to construct a positive series $\sum p_k$ for which $p_k(a_k/a_{k+1}) - p_{k+1} \geq 1$.

Since $\sum a_k$ is convergent, it converges to a real number M , $M = \sum_{i=1}^{\infty} a_i$. Let $p_k = (M - \sum_{i=1}^k a_i)/a_k$. It is easily seen that $p_k > 0$ for any positive integer k and

$$\begin{aligned} p_k(a_k/a_{k+1}) - p_{k+1} &= \left(M - \sum_{i=1}^k a_i \right) / a_{k+1} - \left(M - \sum_{i=1}^{k+1} a_i \right) / a_{k+1} \\ &= a_{k+1}/a_{k+1} = 1. \end{aligned}$$

(2) Suppose that $\sum a_k$ is a divergent positive series. Let $p_k = \sum_{i=1}^k a_i/a_k$. It is easily seen that $p_k > 0$ and

$$p_k(a_k/a_{k+1}) - p_{k+1} = \sum_{i=1}^k a_i/a_{k+1} - \sum_{i=1}^{k+1} a_i/a_{k+1} = -a_{k+1}/a_{k+1} = -1 \leq 0.$$

Now we prove that $\sum 1/p_k$ diverges. We are going to show that for any given positive integer m , there is an $n > m$ such that $\sum_{k=m}^n 1/p_k > 1/2$.

Since $\sum a_k$ diverges and $a_k > 0$, for any given positive integer m , we can always find an $n > m$ such that $a_m + \dots + a_n > a_1 + \dots + a_{m-1}$. Hence

$$\begin{aligned} \sum_{k=m}^n 1/p_k &= a_m/(a_1 + \dots + a_m) + \dots + a_n/(a_1 + \dots + a_n) \\ &> (a_m + \dots + a_n)/(a_1 + \dots + a_n) \\ &= 1/(1 + (a_1 + \dots + a_{m-1})/(a_m + \dots + a_n)) \\ &> 1/(1 + 1) = 1/2. \end{aligned}$$

Therefore $\sum 1/p_k$ diverges by Cauchy's criterion.

ACKNOWLEDGMENT. The author thanks the referee for comments. The author also thanks Professor Scott Hochwald for his efforts to improve this note.

REFERENCES

1. E. Bishop and D. Bridges, *Constructive Analysis*, Springer-Verlag, Berlin, 1985.
2. L. Brand, *Advanced Calculus*, John Wiley & Sons, Inc., New York, 1955.
3. T. J. I. Bromwich, *Introduction to the Theory of Infinite Series*, 2nd edition, MacMillan, New York, 1965.
4. H. Dörrie, *Unendliche Reihen*, Verlag von R. Oldenbourg, München, 1951.
5. G. Faber, *Gesammelte mathematische Abhandlungen*, B. G. Teubner, Leipzig, 1910.
6. M. Fichtenholz, *Infinite Series*, Gordon and Breach Science Publishers Inc., New York, 1970.
7. A. Fridy, *Introductory Analysis*, Harcourt Brace Janovich, Inc., Orlando, 1987.
8. W. Fulks, *Advanced Calculus*, 2nd ed., John Wiley & Sons, Inc., New York, 1969.

9. K. Knopp, *Theory and Application of Infinite Series*, Hafner Publishing Company, New York, 1928.
10. D. R. Lick, *The Advanced Calculus of One Variable*, Meredith Corporation, New York, 1971.
11. T. McCullough and K. Phillips, *Foundation of Analysis in Complex Plane*, Holt, Rinehart and Winston, Inc., New York, 1973.
12. R. Reiff, *Geschichte der Unendlichen Reien*, Wiesbaden, Dr. Martin Sandig oHG., 1889.
13. S. C. Saxena and S. M. Shah, *Introduction to Real Variable Theory*, Intext Educational Publishers, Scranton, 1972.
14. I. S. Sokolnikoff, *Advanced Calculus*, McGraw-Hill, New York, 1939.

Department of Mathematics and Statistics
University of North Florida
Jacksonville, FL 32224

Isometries of ℓ_p -norm

Chi-Kwong Li and Wasin So

The goal of this note is to provide a short proof of the fact that the isometries of ℓ_p -norm ($p \neq 2$) on \mathbf{R}^n are generalized permutation matrices, those matrices that can be written as a product of a diagonal orthogonal matrix and a permutation matrix. As usual, the ℓ_p -norm on \mathbf{R}^n is defined by $\ell_p(x) = (\sum_{i=1}^n |x_i|^p)^{1/p}$ if $1 \leq p < \infty$, and $\ell_p(x) = \max_{1 \leq i \leq n} |x_i|$ if $p = \infty$. An isometry of ℓ_p -norm is an $n \times n$ matrix A satisfying

$$\ell_p(Ax) = \ell_p(x) \quad \text{for all } x \in \mathbf{R}^n.$$

It is well-known that the isometries of the ℓ_2 -norm are orthogonal matrices. A less well-known fact is the following theorem.

Theorem. *Suppose $1 \leq p \leq \infty$ and $p \neq 2$. An $n \times n$ matrix A is an isometry of ℓ_p -norm if and only if A is a generalized permutation matrix.*

This result can be deduced from stronger statements about more general norms [2] [3] [7] or can be viewed as a special case of its infinite dimensional version [1, p. 119] [5, p. 112]. However, there is a direct proof which requires only a basic fact from the theory of norm on \mathbf{R}^n [4, Ch. 5]: an $n \times n$ matrix A is an isometry of ℓ_p -norm if and only if its transpose A^T is an isometry of ℓ_q -norm, where $1/p + 1/q = 1$.

Proof: The sufficiency is easy to check. For necessity, suppose $A = (a_{ij})$ is an isometry of the ℓ_p -norm. We may assume $1 \leq p < 2$, otherwise consider the ℓ_q -norm and A^T . First consider the case that $p \neq 1$. For $i = 1, \dots, n$, let e_i denote the i th column of the $n \times n$ identity matrix. Then $\ell_p(Ae_i) = 1$ for all $i = 1, \dots, n$. Thus $|a_{ij}| \leq 1$ and $\sum_{i,j=1}^n |a_{ij}|^p = n$. Since A^T is an isometry of the ℓ_q -norm, the same argument gives $\sum_{i,j=1}^n |a_{ij}|^q = n$. Notice that $|a_{ij}|^q \leq |a_{ij}|^p$, and equality holds if and only if $|a_{ij}| = 0$ or 1 . Since every column of A (respectively, A^T) has ℓ_p -norm (respectively, ℓ_q -norm) equal to one, the equality $\sum_{i,j=1}^n |a_{ij}|^p = n = \sum_{i,j=1}^n |a_{ij}|^q$ implies that each row and each column of A has exactly one nonzero

entry whose magnitude equals one, and the result follows. For the case $p = 1$, the above argument also yields $\sum_{i,j=1}^n |a_{ij}| = n = \sum_{i=1}^n \alpha_i$, where $\alpha_i = \max_{1 \leq j \leq n} |a_{ij}|$. Thus each row and hence each column of A has only one nonzero entry with magnitude equal to one, and the result follows. ■

After this note was finished, the authors were informed that R. Mathias [6] had obtained the same proof for the complex case previously and independently.

ACKNOWLEDGMENT. The authors wish to thank Professor R. Horn for bringing the references [6] [7] to their attention.

REFERENCES

1. B. Beauzamy, *Introduction to Banach Spaces and Their Geometry*, North Holland, 1985.
2. S. Chang and C. K. Li, Certain isometries on \mathbf{R}^n , *Linear Algebra and its Applications*, 165 (1992) 251–265.
3. D. Z. Dokovic, C. K. Li and L. Rodman, Isometries of symmetric gauge functions, *Linear and Multilinear Algebra*, 30 (1991) 81–92.
4. R. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
5. J. Lindenstrauss and L. Tzafriri, *Classical Banach Spaces I: sequence spaces*, Springer-Verlag, 1977.
6. R. Mathias, unpublished note.
7. C. Wang, et. al., Structure of p -isometric matrices and rectangular matrices with minimum p -norm condition number, *Linear Algebra and its Applications*, 184 (1993) 261–278.

Department of Mathematics
The College of William and Mary
Williamsburg, VA 23187
ckli@cs.wm.edu

Division of Math & Info Sciences
Sam Houston State University
Huntsville, TX 77341
mth.wso@shsu.edu

A Trace Inequality for Unitary Matrices

Bo-Ying Wang and Fuzhen Zhang

The aim of this note is to present a trace inequality for complex unitary matrices, relating the average of the eigenvalues of each of two unitary matrices to that of their product.

$\text{tr}X$, as usual, denotes the trace of the n -square matrix X , which is equal to the sum of the eigenvalues of X , and $m(X)$ designates the algebraic mean of the eigenvalues of X . We show that for any unitary matrices U and V

$$\sqrt{1 - |m(UV)|^2} \leq \sqrt{1 - |\bar{m}(U)|^2} + \sqrt{1 - |m(V)|^2} \quad (1)$$

with equality if and only if U or V is a unitary scalar matrix.

(1) follows from an inequality in an inner product space V with inner product (\cdot, \cdot) , which is of interest itself.

For any unit vectors u, v and w in V , we claim that

$$\sqrt{1 - |(u, v)|^2} \leq \sqrt{1 - |(u, w)|^2} + \sqrt{1 - |(v, w)|^2} \quad (2)$$

with equality if and only if w is a multiple of u or v .

To prove this, we first notice that any component of w that is orthogonal to the span of u and v plays no role in (2), i.e., we really have a problem in which u and v are arbitrary unit vectors, w is in the span of u and v , and $(w, w) \leq 1$. The case $w = 0$ is trivial. If $w \neq 0$, scaling up w to have length 1 diminishes the right-hand side of (2), so we are done if we can prove inequality (2) for arbitrary unit vectors u , v and w with w in the span of u and v . The case in which u and v are dependent is trivial. Suppose u and v are independent, and let $\{u, z\}$ be an orthonormal basis of $\text{Span}\{u, v\}$, so that $v = \mu u + \lambda z$ and $w = \alpha u + \beta z$ for some complex numbers μ , λ , α and β , then

$$|\lambda|^2 + |\mu|^2 = 1 \quad \text{and} \quad |\alpha|^2 + |\beta|^2 = 1.$$

Now use these relations and the arithmetic-geometric mean inequality, together with the fact $|c| \geq \text{Re}(c)$ for any complex number c , to compute

$$\begin{aligned} |\lambda\beta| &= |\lambda\beta|(|\mu|^2 + |\lambda|^2 + |\alpha|^2 + |\beta|^2)/2 \\ &\geq |\lambda\beta|(|\lambda\beta| + |\alpha\mu|) \\ &= |\lambda\beta|^2 + |\lambda\beta\alpha\mu| \\ &= |\lambda\beta|^2 + |\lambda\bar{\beta}\alpha\bar{\mu}| \\ &\geq |\lambda\beta|^2 + \text{Re}(\lambda\bar{\beta}\alpha\bar{\mu}), \end{aligned}$$

so that $-2|\lambda\beta| \leq -2|\lambda\beta|^2 - 2\text{Re}(\lambda\bar{\beta}\alpha\bar{\mu})$. Thus we have

$$\begin{aligned} (|\lambda| - |\beta|)^2 &= |\lambda|^2 - 2|\lambda\beta| + |\beta|^2 \\ &\leq |\lambda|^2 + |\beta|^2 - 2|\lambda\beta|^2 - 2\text{Re}(\lambda\bar{\beta}\alpha\bar{\mu}) \\ &= |\lambda|^2 + |\beta|^2(1 - |\lambda|^2) - |\lambda\beta|^2 - 2\text{Re}(\lambda\bar{\beta}\alpha\bar{\mu}) \\ &= (1 - |\mu|^2) + |\beta|^2|\mu|^2 - |\lambda\beta|^2 - 2\text{Re}(\lambda\bar{\beta}\alpha\bar{\mu}) \\ &= 1 - |\mu|^2(1 - |\beta|^2) - |\lambda\beta|^2 - 2\text{Re}(\lambda\bar{\beta}\alpha\bar{\mu}) \\ &= 1 - |\mu\alpha|^2 - |\lambda\beta|^2 - 2\text{Re}(\lambda\bar{\beta}\alpha\bar{\mu}) \\ &= 1 - |\alpha\bar{\mu} + \beta\bar{\lambda}|^2. \end{aligned}$$

This gives

$$(|\lambda| - |\beta|) \leq \sqrt{1 - |\alpha\bar{\mu} + \beta\bar{\lambda}|^2}, \quad \text{or} \quad |\lambda| \leq |\beta| + \sqrt{1 - |\alpha\bar{\mu} + \beta\bar{\lambda}|^2},$$

which is the same as

$$\sqrt{1 - |\mu|^2} \leq \sqrt{1 - |\alpha|^2} + \sqrt{1 - |\alpha\bar{\mu} + \beta\bar{\lambda}|^2}.$$

Since $|\mu|^2 = |(u, v)|^2$, $|\alpha|^2 = |(u, w)|^2$, and $|\alpha\bar{\mu} + \beta\bar{\lambda}|^2 = |(\alpha u + \beta z, \mu u + \lambda z)|^2 = |(w, v)|^2$, the inequality (2) is proved.

Equality holds for the overall inequality if and only if equality holds at the two points in our derivation where we invoked the arithmetic-geometric mean inequality and $|c| \geq \operatorname{Re}(c)$. Thus, equality holds if and only if $|\lambda| = |\beta|$ and $|\alpha| = |\mu|$, as well as $\operatorname{Re}(\lambda\beta\alpha\bar{\mu}) = |\lambda\beta\alpha\bar{\mu}|$. The former is equivalent to having $\lambda = e^{i\theta}\beta$ and $\mu = e^{i\phi}\alpha$ for some real numbers θ and ϕ , while the latter is then equivalent to $\operatorname{Re}\{|\alpha\beta|^2(e^{i(\theta-\phi)} - 1)\} = 0$. Thus, $\alpha = 0$, $\beta = 0$, or $e^{i\theta} = e^{i\phi}$, so equality in (2) holds if and only if either w is a multiple of u ($\beta = 0$) or w is a multiple of v ($\alpha = 0$ or $e^{i\theta} = e^{i\phi}$).

It is readily seen that for unit vectors u, v and w in V ,

$$\left| \sqrt{1 - |(u, v)|^2} - \sqrt{1 - |(u, w)|^2} \right| \leq \sqrt{1 - |(v, w)|^2}.$$

(2) can be rewritten as

$$\sqrt{s - \|w\|^2|(u, v)|^2} \leq \sqrt{s - \|v\|^2|(u, w)|^2} + \sqrt{s - \|u\|^2|(v, w)|^2},$$

for any u, v and w in V , where $s = (\|u\| \|v\| \|w\|)^2$.

Now we return to the inequality (1) by considering the vector space of all n -by- n complex matrices with the inner product $(A, B) = \operatorname{tr} B^*A$ for every pair of matrices A and B , where $*$ means the conjugate transpose.

Putting

$$w = \frac{1}{\sqrt{n}}I, \quad u = \frac{1}{\sqrt{n}}V \quad \text{and} \quad v = \frac{1}{\sqrt{n}}U^*$$

in (2) where I is the n -by- n identity matrix, U and V are any n -by- n unitary matrices, we have

$$\sqrt{1 - \left| \frac{1}{n} \operatorname{tr} UV \right|^2} \leq \sqrt{1 - \left| \frac{1}{n} \operatorname{tr} U \right|^2} + \sqrt{1 - \left| \frac{1}{n} \operatorname{tr} V \right|^2},$$

equality occurs if and only if U or V is a unitary scalar matrix, that is

$$\sqrt{1 - |m(UV)|^2} \leq \sqrt{1 - |m(U)|^2} + \sqrt{1 - |m(V)|^2}$$

with equality if and only if U or V is a unitary scalar matrix.

We end this note by recalling the well-known fact that if A and B are close to the identity I in norm, then so is the product AB . (2) gives some information about the spectrum of the product; if the spectra of two unitary matrices are close to 1 in the sense that the mean of the eigenvalues is near 1, then this “closeness” is essentially preserved under products.

ACKNOWLEDGMENT. We thank the referee for helpful suggestions.

*Department of Mathematics
Beijing Normal University
Beijing 100875, China*

*Department of Mathematical Science
Nova University
Fort Lauderdale FL 33314
zhang@polaris.nova.edu*

An Elementary Proof of the Square Summability of the Discrete Hilbert Transform

Loukas Grafakos

We would like to give an elementary proof of Hilbert's inequality

$$\left(\sum_{j \in \mathbf{Z}} \left| \sum_{\substack{n \in \mathbf{Z} \\ n \neq j}} \frac{a_n}{j - n} \right|^2 \right)^{1/2} \leq \pi \left(\sum_{n \in \mathbf{Z}} |a_n|^2 \right)^{1/2}, \quad (1)$$

where the a_n 's are real and square summable, and also prove that π cannot be replaced by any smaller number.

Hilbert first proved a weaker version of inequality (1), where π was replaced by a larger constant. The original proof used trigonometric series and first appeared in Weyl's [W] doctoral dissertation in 1908. Three years later, Schur [S] obtained a proof of (1), showing that π was the best possible constant. In his proof, he used a version of what we today refer to as Schur's Lemma. This proof can be found in the book [HLP]. Many other proofs and generalizations have been given since then.

The purpose of this note is to give an elementary proof of inequality (1). The proof uses convergence of sequences; remarkably, only one inequality $2ab \leq a^2 + b^2$; and the identity

$$\sum_{\substack{n \in \mathbf{Z} \\ n \neq 0}} \frac{1}{n^2} = \frac{\pi^2}{3}.$$

Before we present the proof, we would like to clarify a point about (1). If (a_n) is square summable, it isn't automatic that the left hand-side of (1) converges. Part of the inequality is the statement that the left hand-side of (1) is finite whenever the right hand-side is.

Assume first that (a_n) is compactly supported, i.e. $a_n = 0$ except for finitely many n . We show below that the left hand-side of (1) is finite and we prove the required inequality for such sequences. Expand the square of the left hand-side of (1). All indices m, n, j below run from $-\infty$ to ∞ unless there is some restriction stated. We obtain

$$\begin{aligned} & \sum_j \sum_{n \neq j} \sum_{m \neq j} a_m a_n \frac{1}{(j - n)(j - m)} \\ &= \sum_n \sum_m a_m a_n \sum_{j \neq n, m} \frac{1}{(j - n)(j - m)}. \end{aligned} \quad (2)$$

Two out of the three sums above are over finite sets of indices and the interchange of summations is justified. The sum over all $m = n$ in (2) is clearly equal to

$$\sum_n a_n^2 \sum_{j \neq n} \frac{1}{(j-n)^2} = \frac{\pi^2}{3} \sum_n a_n^2. \quad (3)$$

Assume below that $m \neq n$. We calculate the sum over j in (2). We have

$$\begin{aligned} \sum_{j \neq m, n} \frac{1}{(j-n)(j-m)} &= \frac{1}{m-n} \sum_{j \neq m, n} \left(\frac{1}{j-m} - \frac{1}{j-n} \right) \\ &= \frac{1}{m-n} \lim_{k \rightarrow \infty} \sum_{\substack{j \neq m, n \\ |j| \leq k}} \left(\frac{1}{j-m} - \frac{1}{j-n} \right) \\ &= \frac{1}{m-n} \lim_{k \rightarrow \infty} \left[\left(\sum_{\substack{j \neq m \\ |j| \leq k}} \frac{1}{j-m} \right) - \frac{1}{n-m} \right. \\ &\quad \left. - \left(\sum_{\substack{j \neq n \\ |j| \leq k}} \frac{1}{j-n} \right) + \frac{1}{m-n} \right] \\ &= \frac{2}{(m-n)^2} + \frac{1}{m-n} \lim_{k \rightarrow \infty} \left[\sum_{\substack{j \neq m \\ |j| \leq k}} \frac{1}{j-m} - \sum_{\substack{j \neq n \\ |j| \leq k}} \frac{1}{j-n} \right] \\ &= \frac{2}{(m-n)^2}, \end{aligned} \quad (4)$$

since the expression inside brackets above has limit 0 as $k \rightarrow \infty$. Because of (4), the off-diagonal terms in (2) are exactly equal to:

$$\sum_n \sum_{m \neq n} a_n a_m \frac{2}{(m-n)^2}. \quad (5)$$

Using the inequality $2a_m a_n \leq a_n^2 + a_m^2$ we bound (5) by:

$$\sum_n \sum_{m \neq n} \frac{a_n^2}{(m-n)^2} + \sum_m \sum_{n \neq m} \frac{a_m^2}{(m-n)^2} = \frac{\pi^2}{3} \sum_n a_n^2 + \frac{\pi^2}{3} \sum_m a_m^2 = \frac{2\pi^2}{3} \sum_n a_n^2. \quad (6)$$

Combining (3) and the estimate (6) for (5), we obtain inequality (1) for compactly supported sequences. A simple limiting argument gives (1) for general square summable sequences.

We now turn to the fact that π is the best possible constant. We define b_N to be (5) divided by $\sum_n a_n^2$, where (a_n) is the sequence 1 for $|n| \leq N$ and 0 otherwise.

Estimate (6) shows that $b_N \leq 2\pi^2/3$. A simple calculation gives

$$b_N \geq \frac{4N}{2N+1} \left[\sum_{k=1}^{N+1} \frac{1}{k^2} \right] + \frac{4(N-1)}{2N+1} \\ \times \left[\frac{\frac{1}{1^2} + \left(\frac{1}{1^2} + \frac{1}{2^2} \right) + \cdots + \left(\frac{1}{1^2} + \frac{1}{2^2} + \cdots + \frac{1}{(N-1)^2} \right)}{N-1} \right].$$

Applying the squeeze law, we obtain that b_N tends to $2\pi^2/3$ as $N \rightarrow \infty$. Using (3) and (5), we obtain that for this choice of (a_n) , the ratio of the left-hand side of (1) and $(\sum_n a_n^2)^{1/2}$ converges to π as $N \rightarrow \infty$. This proves that π is the best possible constant in (1), Q.E.D.

Inequality (1) is known to be strict if (a_n) is nonzero. To see this for compactly supported nonzero sequences, observe that (6) is a strict bound for (5) since for some m and n , $2a_n a_m < a_n^2 + a_m^2$. For general sequences, a further argument is needed since the passage to the limit will destroy the strict inequality. See [HLP] for details on this.

To the best of my knowledge, the determination of the best constant for the l^p inequality, $1 < p \neq 2 < \infty$ remains unresolved. Pichorides [P] solves this problem for the corresponding continuous operator.

We end this note by asking a question: Is there a constant C such that for all square summable sequences (a_n) and all bounded sequences (λ_n) , the following inequality holds?

$$\left(\sum_{j \in \mathbf{Z}} \left| \sum_{\substack{n \in \mathbf{Z} \\ n \neq j}} \lambda_{j+n} \frac{a_n}{j-n} \right|^2 \right)^{1/2} \leq C \sup_j |\lambda_j| \left(\sum_{n \in \mathbf{Z}} |a_n|^2 \right)^{1/2}. \quad (7)$$

If $\lambda_n = 1$ for all n , then (7) reduces to (1) and one can take $C \geq \pi$.

REFERENCES

- [HLP] G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, Cambridge University Press (1934).
- [P] S. Pichorides, *On the Best Values of the Constants in the Theorems of M. Riesz, Zygmund and Kolmogorov*, *Studia Mathematica* 46 (1972), 164–179.
- [S] I. Schur, *Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen*, *Journal f. Math.* 140 (1911), 1–28.
- [W] H. Weyl, *Singuläre Integralgleichungen mit besonderer Berücksichtigung des Fourierschen Integraltheorems*, Doctoral Dissertation, University of Göttingen (1908).

Department of Mathematics
Washington University, Campus Box 1146
One Brookings Drive, St Louis, MO 63130-4899
grafakos@math.wustl.edu

Answer to Picture Puzzle
(p. 449)
Henri Cartan, the son of Eli Cartan.

THE EVOLUTION OF . . .

Edited by Abe Shenitzer

Mathematics, York University, North York, Ontario M3J 1P3, Canada

How Hyperbolic Geometry Became Respectable

Abe Shenitzer

Many accounts of the evolution of hyperbolic geometry mention the names of Beltrami and Klein but give few details of their contributions in this area. The present paper focuses on some of these very details. The reader is assumed to have some familiarity with basic concepts of differential geometry.

HISTORICAL INTRODUCTION. The discovery of hyperbolic geometry* led to the realization—at the end of the 19th century—that the truths of mathematics are relative rather than absolute, and to the resolution of the millennial doubts about Euclid’s parallel postulate. J. Coolidge describes its effect as follows:

The point which I wish to insist on . . . is that it is to the doubts about Euclid’s parallel postulate, and efforts of such thinkers as Saccheri, Lobachevski, Bolyai, Beltrami, Riemann and Pasch to settle these doubts, that we owe the whole modern abstract conception of mathematical science. ([1], p. 87.)

Another way of assigning to the discovery of hyperbolic geometry its due place is to note that it was one of the two greatest geometric discoveries (or inventions, if you are not a Platonist) of the 19th century, the other being the discovery (by Riemann) of the concept of an n -dimensional manifold. Of course, while the first of these discoveries marked a profound intellectual discontinuity, the second did not; in fact, one can safely say that an n -dimensional manifold and its geometry are direct “descendants” of the notions of a surface and of Gauss’ notion of the intrinsic geometry of a surface.

The story of the discovery of hyperbolic geometry can be conveniently divided into three parts. The first was largely negative in the sense that it consisted of doomed attempts to deduce the Euclidean parallel postulate from the other postulates of Euclidean geometry. The first of these attempts was presumably due

*The terms “hyperbolic geometry” and “hyperbolic plane” refer to the system of plane geometry based on Euclid’s axioms with his parallel axiom replaced by the “hyperbolic parallel axiom” which asserts the existence of a line a and a point A not on a such that there are at least two lines passing through A that are parallel to a . An isometric replica of the hyperbolic plane is called a model of that plane.

to Archimedes. Other such attempts were due to Arab mathematicians, and, more recently, to Saccheri (1733) and Legendre (1794).

The second part of our story involved attempts in the first third of the 19th century to deduce basic logical consequences of the axioms of Euclidean geometry with its parallel axiom replaced by the hyperbolic parallel axiom. The mathematicians involved were Schweikart, Lambert, and, more importantly, Gauss, Bolyai, and Lobachevski. Gauss and Lobachevski considered the possibility that the geometry of the universe is hyperbolic rather than Euclidean. Bolyai's chief concern was about the consistency of the new geometry. This is where matters stood in the 1830s. During the next 30 years the effect on mathematics of these investigations and insights was virtually nil.

The third stage of the evolution of hyperbolic geometry began in the 1860s and ended in the 1880s. Here the key contributions—the first models of the hyperbolic plane—were due to Beltrami and, in part, to Klein. Their work provided a proof of the relative consistency of hyperbolic geometry and put it logically on a par with Euclidean geometry.

TECHNICAL ACCOUNT. What follows is a description, in four parts, of the genesis and nature of the early models of the hyperbolic plane due to Beltrami and, in part, to Klein. The approach in the first part is based on an essay by A. M. Lopshits in volume II of [4].

(α) Beltrami studied mappings of surfaces in E^3 into the Euclidean plane that preserve geodesics, that is, map geodesics to (straight) lines in the plane.

A familiar example of such a mapping is the central projection of a sphere to the plane. We give a few highly relevant details of this mapping.

Consider (FIGURE 1) the projection of the lower hemisphere onto the plane. If $m(X, Y)$ is the plane image of M on the sphere, then we assign (X, Y) to M as its curvilinear coordinates. If the radius of the sphere is a , then its (Gaussian) curvature is $1/a^2$ and its length element ds^2 turns out to be

$$ds^2 = \frac{(1 + KY^2) dX^2 + (1 + KX^2) dY^2 - 2KXY dX dY}{[1 + K(X^2 + Y^2)]^2}, \quad (1)$$

with K the (Gaussian) curvature.

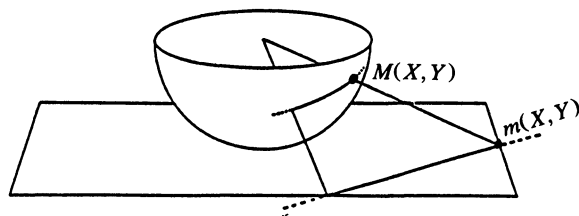


Figure 1

In 1866 Beltrami showed, quite generally, that if a surface in E^3 can be mapped in a one-one geodesic manner into the plane, then, after transfer of the coordinates (X, Y) of the image points to the points of the surface, its length element ds^2 can be reduced to the form (1) with K its constant curvature. Furthermore, the only surfaces that admit such mappings are surfaces with constant curvature. Since surfaces with the same ds^2 are locally isometric, it follows that a surface with

constant curvature $K > 0$ is locally isometric to a sphere, and a surface with constant curvature $K = 0$ is locally isometric to a plane. The case of greatest interest to Beltrami was that of pseudospherical surfaces, that is, surfaces in E^3 with constant curvature $K < 0$. Beltrami studied such surfaces in “Saggio...”, the first of the two papers he published in 1868. What follows is a brief summary of his findings.

In (1) put $K = -1/k^2$, $X = kx$, and $Y = ky$. Then the length element of a pseudospherical surface takes the form

$$ds^2 = k^2 \frac{(1 - y^2) dx^2 + (1 - x^2) dy^2 + 2xy dx dy}{(1 - x^2 - y^2)^2}, \quad (2)$$

where k is a positive constant. Since the necessarily positive discriminant of the form (2) is equal to

$$k^4 / (1 - x^2 - y^2)^3, \quad (3)$$

it follows that

$$x^2 + y^2 < 1.$$

This means that a geodesic mapping of a pseudospherical surface into a Euclidean plane carries each point of the surface to a point of the unit disk (see FIGURE 2).

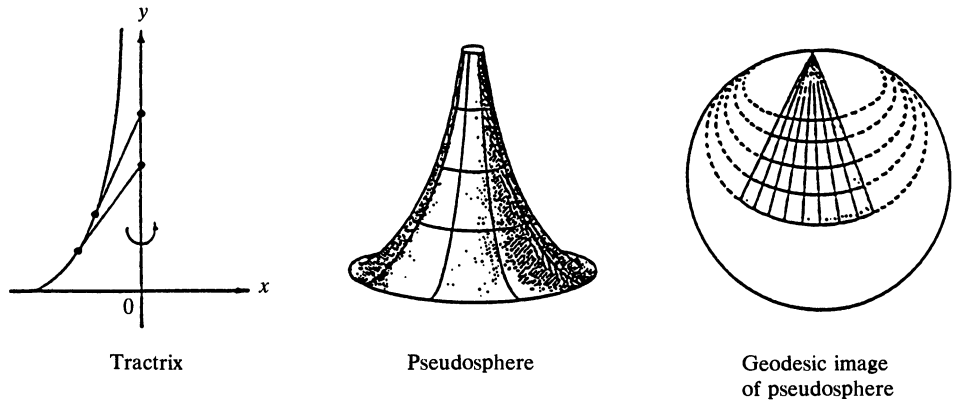


Figure 2

Now there is a 1-1 geodesic mapping of the hyperbolic plane onto the unit disk (see [2]. Also see FIGURE 3; the “dish” represents a unit disk that is part of a horosphere tangent to a hyperbolic plane. The curved lines represent hyperbolic straight lines belonging to the mapping pencil of parallel lines). If (as in the case of

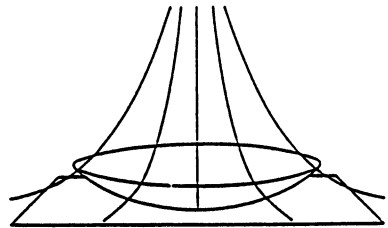


Figure 3

the stereographic projection of the hemisphere) we transfer the coordinates (x, y) of a point in the unit disk to its preimage in the hyperbolic plane, then the length element ds^2 of the latter takes the form (2) *The sameness of the length elements of the pseudospherical surface and of the hyperbolic plane proves that these two surfaces are locally isometric.*

Beltrami's realization of the geometry of (part of) the hyperbolic plane as the intrinsic geometry of a pseudospherical surface had a tremendous impact on his contemporaries. It was seen as changing what was thought to be a figment of the imagination into a mathematical fact.

If a pseudospherical surface is to be an isometric replica of the full hyperbolic plane, then it must have the basic property that the geodesics through any of its points are infinite in both directions. This turns out to be equivalent to the assumption that there exists a pseudospherical surface that admits a one-one geodesic mapping *onto* the unit disk. Beltrami thought that this is a defensible assumption. It is not, at least for smooth surfaces in E^3 . In 1902 Hilbert showed that there is no such surface in E^3 that is an isometric replica of the hyperbolic plane.

(β) While the main emphasis in Beltrami's "Saggio..." is on showing that a pseudospherical surface is locally isometric to the hyperbolic plane, it also points to the possibility of turning the unit disk into a model of the hyperbolic plane by exploiting various features of the geodesic mapping linking it to a pseudospherical surface. Hence the term "the Beltrami disk model." The relevant definitions are: a hyperbolic (straight) line is a chord of the unit disk, hyperbolic parallel lines are chords that meet at a point of the unit circle, and hyperbolic diverging lines are chords that don't meet in the closed unit disk. The source of quantitative relations is the length element ds^2 given by equation (2). (Thus the length of a curve $x = x(t)$, $y = y(t)$ is $\int ds$, with suitable limits in the integral.) The key missing element in Beltrami's disk model is the definition of a motion.

(γ) In 1868 Dedekind published Riemann's inaugural lecture of 1854. After reading it, Beltrami published the second of his 1868 papers in which he once more obtained his earlier disk model as well as *three conformal models* of the hyperbolic plane. The following description of these models is taken from the introduction in J. Stillwell's translation of the second of Beltrami's 1868 papers.

The starting point is a *hemisphere model* consisting of the open hemisphere

$$x_1^2 + x_2^2 + x^2 = a^2, \quad x > 0 \quad (4)$$

in 3-dimensional (x_1, x_2, x) -space, with the line element

$$ds = R \frac{\sqrt{dx_1^2 + dx_2^2 + dx^2}}{x}. \quad (5)$$

The geodesics for this metric are vertical sections of the hemisphere. Perpendicular projection of this model onto the plane $x = 0$ which we take to be horizontal (FIGURE 4) yields Beltrami's earlier disk model.

The third model is obtained by stereographic projection of the hemisphere onto its horizontal tangent plane (FIGURE 5), and the line element is computed to be

$$ds = \frac{\sqrt{d\xi_1^2 + d\xi_2^2}}{1 - \frac{1}{4R^2}(\xi_1^2 + \xi_2^2)}, \quad (6)$$

a metric which was stated by Riemann (1854) to be of constant curvature. We

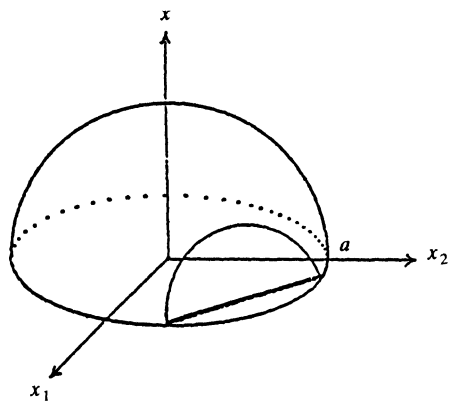


Figure 4

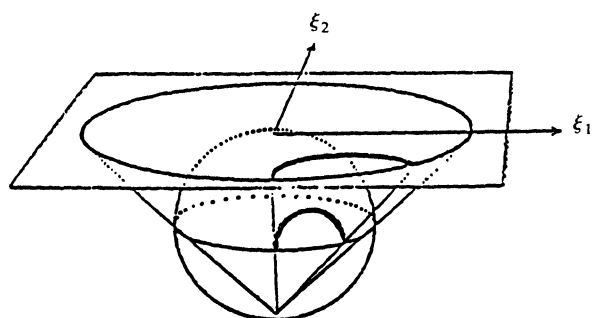


Figure 5

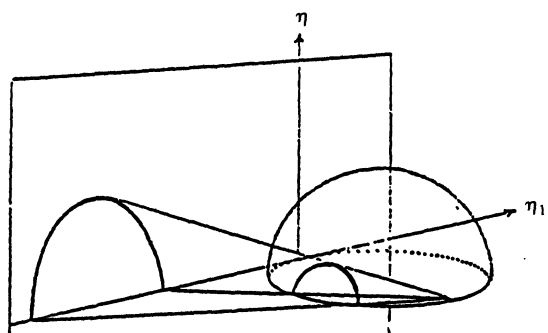


Figure 6

observe that (6) is a multiple of the Euclidean line element $\sqrt{d\xi_1^2 + d\xi_2^2}$ in the (ξ_1, ξ_2) -plane; the multiple of course varies with position but not with direction, hence angles are preserved. Thus we have the conformal, or “Poincaré disk” model, 14 years before its appearance in Poincaré (1882).

The fourth model is obtained by stereographic projection of the hemisphere onto (the upper half of) a vertical plane (FIGURE 6). This gives the line element

$$ds = R \frac{\sqrt{d\eta_1^2 + d\eta_2^2}}{\eta} \quad (7)$$

for the “Poincaré half plane.”

(δ) In 1871 Klein used the unit disk and its chords to create his so-called projective model of the hyperbolic plane. His points, lines, parallel lines, and diverging lines were the same as Beltrami’s but he based his definitions of distance, angle, and motion on the ideas of Cayley. He defined the distance between points M_1 and M_2 (see FIGURE 7) as, essentially, the logarithm of the cross ratio $\{P_1P_2, M_2M_1\}$,

$$M_1M_2 = \frac{k}{2} \ln J = \frac{k}{2} \ln \left(\frac{P_1M_2}{P_1M_1} \middle/ \frac{P_2M_2}{P_2M_1} \right),$$

defined the angle between two lines by, essentially, a projective dual of his definition of distance, and defined a motion of his model as a projective transformation of the plane of the disk that mapped its boundary onto itself. Klein’s version of the unit-disk model of the hyperbolic plane was the first instance of a geometry in the sense of his famous Erlangen Program of 1872.

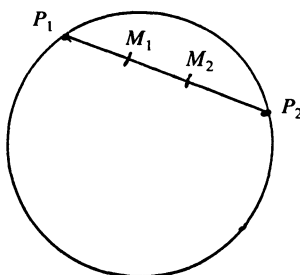


Figure 7

In sum we can say that the works of Beltrami and Klein that appeared between 1866 and 1871 put hyperbolic geometry on a par with Euclidean geometry. *From then on it became ever clearer that the truths of mathematics are relative rather than absolute.*

REFERENCES

1. J. L. Coolidge, A history of geometrical methods, Dover, 1963. Chapter IV, pp. 68–87. (“Nice summary”)
2. N. V. Efimov, Higher geometry, Mir, 1980. Chapter 8, pp. 474–528. (Useful beyond the needs of this essay.)
3. H. Eves, Great moments in mathematics, DME #7, MAA, 1983. Lectures 26 and 27. (Fine elementary account.)
4. V. F. Kagan, Foundations of geometry, GITTL, Moscow, Parts I and II. (In Russian.) Lectures 26 and 27. (Included because Part II contains the Lopshits essay mentioned on p. 465.)
5. J. Stillwell, Mathematics and its history, Springer, 1989. Chapter 17, pp. 255–274.
6. I. M. Yaglom, Geometric transformations III, NML #24, MAA, 1973. Introduction and Supplement, pp. 103–135. (Supplement contains detailed elementary account of Klein’s disk model.)

*Department of Mathematics
York University
North York, Ontario
CANADA M3J 1P3*

The Spirit is Willing but the Ham is Rotten

John Kinloch and Rick Norwood

What is wrong with the following “proof” of the Ham Sandwich Theorem?

Theorem: Given any sandwich composed of bread, ham, and cheese, there is a single plane which cuts the sandwich into two parts, such that the two parts contain equal amounts of bread, equal amounts of ham, and equal amounts of cheese.

Proof: Consider the bread. It has a center of mass. Call this point p . Let q be the center of mass of the ham. Let r be the center of mass of the cheese. There is a plane containing p , q , and r . This plane divides the sandwich into two parts containing equal amounts of bread, equal amounts of ham, and equal amounts of cheese.

The theorem is true. The “proof” is fallacious. Why?

*Department of Mathematics
East Tennessee State University
Johnson City TN 37614*

THE AUTHORS

THOMAS BANCHOFF received his B.A. from Notre Dame in 1960 and his Ph.D. from Berkeley in 1964. He taught as a Benjamin Peirce Instructor at Harvard and a Research Associate at the University of Amsterdam before coming in 1967 to Brown University where he has been ever since. Most of his research and teaching has focused on geometry of higher dimensions and visualization by means of computer graphics, and he particularly likes to investigate the interplay between smooth and polyhedral phenomena. In 1978 with Charles Strauss, he produced the computer-generated film *The Hypercube: Projections and Slicing*. His recent works include the Scientific American Library volume *Beyond the Third Dimension* and a new introduction for this favorite book *Flatland*.

PETER GIBLIN did his graduate work at London University and joined the faculty at Liverpool University in 1967, where he is now a Reader, but with little leisure in which to read. He was a visiting professor at the University of North Carolina at Chapel Hill in 1981/2 and at the University of Massachusetts at Amherst in 1985/6, where he was the first Five College Professor of Geometry (a title to impress even an Englishman). It was during this period that he first visited Brown University and started a fruitful collaboration with Tom Banchoff. He is the co-author of the book *Curves and Singularities*, with J. W. Bruce, who coincidentally was the second Five College Professor of Geometry. His research interests are in the applications of singularity theory to differential geometry and computer vision.

ELLIOT LINZER received B.S. degrees in electrical engineering and mathematics from the University of Maryland in 1987. In 1990 he received a Ph.D. in electrical engineering from Columbia University. He then joined IBM Research; first in the mathematical sciences department and now in computer science. His research interests include numerical algorithms, numerical analysis, signal processing and image compression.

ALAIN ROBERT is a Professor of Mathematics at Neuchâtel University (CH) since 1971. He received his Ph.D. from Neuchâtel (1967), visited Paris (67-68), Princeton Univ. (68-70) and IAS (70-71), Queen's Univ. (Kingston, Ont. 73-74) and Berkeley (83-84). He is the author of *Elliptic Curves*, *Group Representation*, *Nonstandard Analysis*, and *Advanced Calculus*. He is currently involved in research in p -adic analysis and congruence properties of special polynomials.

RICHARD P. SAVAGE, JR. received his Ph.D. in mathematics in 1981 at the University of Utah under the direction of Domingo Toledo. In 1982 he joined the mathematics department at Tennessee Technological University. His main research interests are in differential geometry. Besides mathematics, he is interested in genealogy and in spending as much time as possible outdoors.

LAURENT BEECKMANS, born in 1967, studied Mathematics at the Université Libre de Bruxelles, Belgium, where he completed his graduate work on Egyptian Fractions in 1989 under the direction of Professor Jean Doyen. In 1991-92 he worked with Professor Graham Everest at the University of East Anglia, England. His main research interests are in Elementary Number Theory and Discrete Mathematics. Some of his numerous hobbies are constructing models of polyhedrons, studying languages, collecting cans, playing and composing music.

STEPHEN TURNER, a recent emarkee upon a number theory Ph.D. at Glasgow, and whose experience of forming knots as a seaman did not conform to his topological intuition, studied at Liverpool and Cambridge Universities. Having attended an earlier degree course until he stopped—a

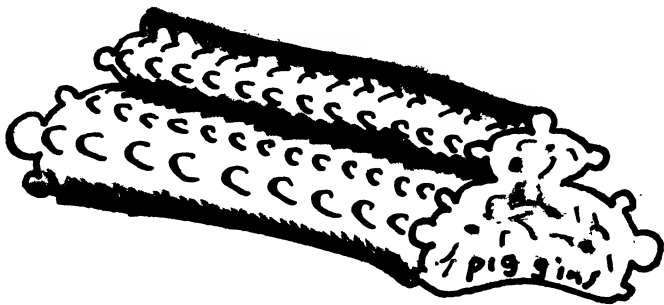
decision in which the institution concurred—he worked and dabbled in math, later going on to study part-time with Britain’s Open University, and then full-time as above.

The present article is based on an undergraduate project at Liverpool.

PAUL T. BATEMAN obtained three degrees from the University of Pennsylvania, writing his Ph.D. thesis in 1946 under Hans Rademacher. From 1950 until his retirement in 1989 he was at the University of Illinois in Urbana-Champaign. He was a coeditor of the Problems Section of the MONTHLY from 1986 through 1991. His mathematical interests center around number theory; other interests include opera, classical music, spectator sports, driving the back roads of Colorado, and swimming badly.

HAROLD G. DIAMOND received his doctorate from Stanford University in 1965, and has been at the University of Illinois since 1967. He describes himself as a first-digit number theorist, interested in matters such as sieves and asymptotic estimates for prime numbers. This taste was acquired through contact with Mark Kac, Paul Turán, K. Chandrasekharan and his thesis supervisor, Paul J. Cohen. He is a problems enthusiast and served as a coeditor of the MONTHLY Problems Section. His personal interests include photography.

ANITA E. SOLOW received a B.A. in mathematics from the University of Pennsylvania and a Ph.D. from Dartmouth College. She is Professor of Mathematics at Grinnell College, where she has been on the faculty since 1980. For the past six years, she has been working on calculus reform and is the editor of the recently published book, *Learning by Discovery*. She also is the proud mother of two sons.



MANDELBROT

PROBLEMS AND SOLUTIONS

Edited by:
Richard T. Bumby, Fred Kochman and Douglas B. West

Proposed problems should be sent to the MONTHLY PROBLEMS address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.

Solutions of published problems should arrive before October 31, 1994 at the MONTHLY PROBLEMS address given on the inside front cover. Solutions should be typed with double spacing, including the problem number and the solver's name and mailing address. Two copies suffice. A self-addressed postcard or label should be included if an acknowledgment is desired.

*An asterisk (*) after the number of a problem, or part of a problem, indicates that no solution is currently available. Partial solutions will be useful in such cases. Otherwise, the published solution is likely to be based on a solution which is complete and correct. Of course, an elegant partial solution or a method leading to a more general result is always useful and welcome. In addition, references to other appearances of MONTHLY problems or to solutions of these problems in the literature are also solicited.*

PROBLEMS

10382. *Proposed by Richard K. Guy, University of Calgary, Calgary, Alberta, Canada.*

Which integers are represented by

$$\frac{(x + y + z)^2}{xyz}$$

where x , y , and z are positive integers?

10383. *Proposed by Kevin Ford (student), University of Illinois, Urbana, IL.*

Let B_1, B_2, \dots, B_s denote subsets of a finite set B , and let $\lambda_i = \#(B_i)/\#(B)$ and $\lambda = \lambda_1 + \dots + \lambda_s$. Show that, for every integer t satisfying $1 \leq t \leq \lambda$, there exist r_1, r_2, \dots, r_t with $r_1 < r_2 < \dots < r_t$ and

$$\#(B_{r_1} \cap B_{r_2} \cap \dots \cap B_{r_t}) \geq (\lambda - t + 1) \binom{s}{t}^{-1} \#(B).$$

10384. *Proposed by Franklin Kemp, East Texas State University, Commerce, TX.*

Suppose $x_1 < x_2 < \dots < x_n$ and $y_1 < y_2 < \dots < y_n$. Define the correlation coefficient r in the usual way:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \cdot \sum_i (y_i - \bar{y})^2}}$$

where \bar{x} and \bar{y} are the average values of the x_i and y_i , respectively, and the sums run from 1 to n . Show that $r \geq 1/(n-1)$.

10385. *Proposed by Nándor Sieben, Arizona State University, Tempe, AZ.*

Let X be a topological space. It is easy to see that if X is a Hausdorff space, then *fixed-point sets are closed*. That is, for any continuous function $f: X \rightarrow X$, the set $F = \{x \in X: f(x) = x\}$ is closed. Is the converse true? More precisely, if X has the property that all fixed-point sets are closed, must X be a Hausdorff space?

10386. *Proposed by Jordan Tabov, Bulgarian Academy of Sciences, Sofia, Bulgaria.*

Let a tetrahedron with vertices A_1, A_2, A_3, A_4 have altitudes that meet in a point H . For any point P , let P_1, P_2, P_3 and P_4 be the feet of the perpendiculars from P to the faces $A_2A_3A_4, A_3A_4A_1, A_4A_1A_2$ and $A_1A_2A_3$, respectively. Prove that there exist constants a_1, a_2, a_3 and a_4 such that one has

$$a_1 \overrightarrow{PP_1} + a_2 \overrightarrow{PP_2} + a_3 \overrightarrow{PP_3} + a_4 \overrightarrow{PP_4} = \overrightarrow{PH}$$

for all points P .

10387*. *Proposed by Stanley Rabinowitz, Westford, MA, and Peter J. Costa, University of St. Thomas, St. Paul, MN.*

Let $T_n = (t_{i,j})$ be the n by n matrix with $t_{i,j} = \tan(i+j-1)x$, i.e.,

$$T_n = \begin{pmatrix} \tan x & \tan 2x & \tan 3x & \dots & \tan nx \\ \tan 2x & \tan 3x & \tan 4x & \dots & \tan(n+1)x \\ \tan 3x & \tan 4x & \tan 5x & \dots & \tan(n+2)x \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tan nx & \tan(n+1)x & \tan(n+2)x & \dots & \tan(2n-1)x \end{pmatrix}.$$

Computer experiments suggest that

$$\det(T_n) = (-1)^{\lfloor n/2 \rfloor} \sec^n nx \prod_{r=1}^{n-1} (\sin^2(n-r)x \sec rx \sec(2n-r)x)^r \\ \times \begin{cases} \sin n^2 x & \text{if } n \text{ odd,} \\ \cos n^2 x & \text{if } n \text{ even.} \end{cases}$$

Prove or disprove this conjecture.

10388. *Proposed by E. Sparre Andersen and Mogens Esrom Larsen, Københavns Universitet, København, Denmark.*

Find

$$\sum_{k=0}^n \binom{n}{k} \left(\frac{n-3}{4} - \frac{k}{2} + p \right)$$

where n and p are positive integers.

NOTES

Notes (10386) Although the altitudes of any triangle are concurrent, only special *tetrahedra* will have altitudes that meet in a point, as hypothesized in this problem. A characterization of such tetrahedra might be a useful supplement to the stated problem. Also note that the equation to be proved is a *vector* equation. **(10387)** The determinant of T_n tests whether the sequence $\langle \tan nx \rangle$ satisfies a linear recurrence. The conjectured value of $\det(T_n)$ has been verified symbolically for $n \leq 5$, and tested numerically for several values of x for $6 \leq n \leq 20$.

SOLUTIONS

Some Integrals With a Common Value

10241 [1992, 675]. *Proposed by Roger W. Johnson, Carleton College, Northfield, MN.*

Let m and n be positive integers with $m \geq n$. Show that

$$\int_0^\infty \left(\frac{\sin x}{x} \right)^n \left(\frac{\sin(mx)}{x} \right) dx = \frac{\pi}{2}.$$

Solution I by C. Georgiou, University of Patras, Patras, Greece. Let $I(m, n)$ denote the given integral. Then an integration by parts gives

$$\begin{aligned} \int \left(\frac{\sin x}{x} \right)^n \left(\frac{\sin(mx)}{x} \right) dx &= \int (\sin x)^n \sin(mx) d\left(\frac{-1}{nx^n} \right) \\ &= - \frac{\sin^n x \sin(mx)}{nx^n} \\ &\quad + \frac{1}{n} \int \frac{n \sin^{n-1} x \cos x \sin(mx) + m \sin^n x \cos(mx)}{x^n} dx \end{aligned}$$

from which we get for $n \geq 1$

$$2nI(m, n) = (n + m)I(m + 1, n - 1) + (n - m)I(m - 1, n - 1). \quad (1)$$

Note also that $I(0, n) = 0$ for $n \geq 0$ and $I(m, 0) = \pi/2$ for $m \geq 1$ ($I(m, 0)$ is essentially the standard Dirichlet integral.)

The result follows by an easy induction using (1). Indeed, for $n = 1$ and $m \geq 1$, (1) gives $I(m, 1) = \pi/2$ for $m \geq 1$. Suppose that for $n = k$ ($k \geq 1$) and $m \geq k$ we have $I(m, k) = \pi/2$. Then it follows from (1) that for $m \geq k + 1$, $I(m, k + 1) = \pi/2$, and the induction step is completed.

Solution II by Donald A. Darling, Newport Beach, CA. Let $S_n = X_1 + X_2 + \cdots + X_n$ be the sum of n independent random variables each uniformly distributed over the interval $[-1, 1]$. Each X_i has the characteristic function $E(e^{itX}) = \sin t/t$ and since S_n has a continuous distribution function the inversion formula yields

$$\begin{aligned} \mathbf{P}(|S_n| \leq m) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{\sin t}{t} \right)^n \int_{-m}^m e^{-it\xi} d\xi dt \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \left(\frac{\sin t}{t} \right)^n \left(\frac{\sin mt}{t} \right) dt \\ &= \frac{2}{\pi} \int_0^{\infty} \left(\frac{\sin t}{t} \right)^n \left(\frac{\sin mt}{t} \right) dt. \end{aligned}$$

If $m \geq n$ this probability is 1, yielding the required expression. It is not necessary that m be an integer.

Editorial comment. Readers employed a variety of other methods including complex contours and self-contained Fourier arguments along the lines of Solution II. The integral can also be found as item number 3.836.1 in I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*, prepared by A. Jeffrey, Academic Press, 1980. Many readers noted that m need not be an integer.

Solved also by F. Alouges & R. Cerf (France), K. F. Andersen (Canada), J. Andraos (Canada), J. Anglesio (France), P. R. Arner & T. R. Falconieri, S.-J. Bang (Korea), R. J. Bass & S. Byrd, K. L. Bernstein, D. Borwein & P. Borwein (Canada), P. Bracken (Canada), J. L. Brown Jr., R. J. Chapman (U. K.), Y. Diao, R. Drnovšek (student, Slovenia), J. Elstrodt (Germany), M. Golomb, J. Hatcher, M. Hoffman, G. L. Isaacs, F.-A. Izadi (Iran), A. S. Izotov (Russia), I. Kastanas, M. S. Klamkin (Canada), O. P. Lossers (The Netherlands), T. L. McCoy, K. McInturff, J. Milcetic, A. Pedersen (Denmark), F. W. Steutel (The Netherlands), D. Tan, N. S. Thorner, M. Vowe (Switzerland), J. Wimp, and the University of Wyoming Problem Circle.

The Maximal Miquel Ratio

10244[1992, 675]. *Proposed by Ken Bromberg (student), Brown University and Stan Wagon, The Geometry Center, Minneapolis, MN and Macalester College, St. Paul, MN.*

A classical construction of Miquel starts with an n -vertex polygon and a point P in the plane (not a vertex of the n -gon), and forms another n -gon as follows:

1. draw the perpendiculars from P to the (extended) sides of the polygon;
2. connect the feet to obtain another n -gon.

These steps are then repeated n times (provided that none of the polygons has P as a vertex). The resulting polygon, denoted $M(P)$ is similar to the initial n -gon.

(a) Given a triangle, construct the point P for which $M(P)$ is largest.

(b)* Given a quadrilateral, is there a Euclidean construction of the point P for which $M(P)$ is largest?

Solution of (a) by the proposers. The solution uses the notation of Figure 10244.

Start with $\triangle ABC$ and assume $A \geq B \geq C$, where A denotes the triangle's angle at A , etc. (Angles are measured so that they lie in the interval $(0, \pi)$.) Then the desired point is the excenter of the triangle opposite A , that is, the intersection of the angle bisector of A with the external angle bisectors at B and C . To prove this, we first show that the similarity constant between $M(P)$ and the initial triangle is $\sin \alpha_1 \sin \alpha_2 \sin \alpha_3$, where α_i (and β_i) are the angles in Figure 10244. Consider the illustrated case, where P is inside $\triangle ABC$. Let $A_1B_1C_1$, $A_2B_2C_2$, $A_3B_3C_3 = M(P)$ be the triangles produced by the Miquel construction. Then $\sin \alpha_1 = PA_1/PA$. Further, because of the cyclic quadrilateral PA_1BB_1 , $\angle PA_1B_1 = \angle PBB_1 = \alpha_2$ and so $\sin \alpha_2 = PA_2/PA_1$. Similarly $\sin \alpha_3 = PA_3/PA_2$. Therefore $\sin \alpha_1 \sin \alpha_2 \sin \alpha_3 = PA_3/PA$, which is the desired similarity constant. A virtually identical argument works when P is outside the triangle. (Incidentally, this permutation of angles proves that $M(P)$ is similar to $\triangle ABC$, for it yields that the angles of $A_3B_3C_3$ are $\alpha_1 + \beta_1, \alpha_2 + \beta_2, \alpha_3 + \beta_3$.)

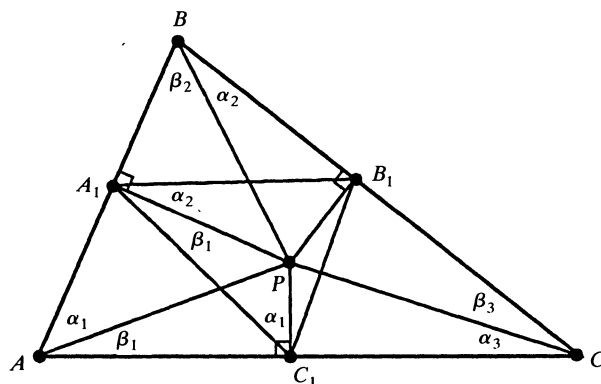


Figure 10244

Now, because $\sin \alpha_1 \sin \alpha_2 \sin \alpha_3 = \sin \beta_1 \sin \beta_2 \sin \beta_3$ (easily proved; it also follows from the preceding argument by symmetry, since $\sin \beta_1 \sin \beta_2 \sin \beta_3$ equals the same similarity constant), it suffices to maximize $f(P) = \sin \alpha_1 \sin \beta_1 \sin \alpha_2 \sin \beta_2 \sin \alpha_3 \sin \beta_3$. But because one of $\alpha_i \pm \beta_i$ is constant (the two cases depend on where P lies within the interior or exterior angle at the vertex), easy calculus shows that the maximum value of $\sin \alpha_i \sin \beta_i$ occurs when $\alpha_i = \beta_i$ or $\alpha_i = \pi - \beta_i$, that is, when P is on one of the internal or external bisectors of the angle at the vertex. So, at vertex B for example, the maximum value of $\sin \alpha_2 \sin \beta_2$ is one of $\sin^2 B/2$ or $\sin^2(\pi - B)/2$ (which equals $\cos^2 B/2$); because B is acute, the maximum is the latter. Similarly, the largest that $\sin \alpha_3 \sin \beta_3$ can be is $\sin^2 C/2$.

Now, the value of $f(P)$ at the excenter opposite A is

$$\sin^2 A/2 \cos^2 B/2 \cos^2 C/2.$$

If A is obtuse we are clearly done, because $\sin^2 A/2$ is then the maximum of $\sin \alpha_1 \sin \beta_1$, so $f(P)$ is certainly the largest it could possibly be. If A is acute, however, then $\sin \alpha_1 \sin \beta_1$ can be larger than $\sin^2 A/2$: if P is external to the angle at A then this value could be as great as $\cos^2 A/2$. But if P is external to A then it can be external to at most one of B and C . It follows that for such P , $f(P) \leq \cos^2 A/2 \cos^2 B/2 \sin^2 C/2$. But this last is less than the value of g at

the excenter opposite A , for that equals $\sin^2 A/2 \cos^2 B/2 \cos^2 C/2$. Thus this excenter maximizes $f(P)$, as claimed. This maximum value is not attained at any other point, since each of the three factors $\sin \alpha_i \sin \beta_i$ is maximized on a bisector of an angle or external angle of the triangle.

Addendum. The maximum Miquel ratio, $\sin A/2 \cos B/2 \cos C/2$, is equal to $s(s-b)(s-c)/(abc)$, where s is the semiperimeter of the triangle and a, b, c are the side-lengths; this is a consequence of some well-known formulas (see Z. A. Melzak, *Invitation to Geometry*, Wiley, 1983). It follows that the maximum Miquel ratio is between $3/8$ and 1 , with the lower bound occurring when the initial triangle is equilateral.

Editorial comment. As the solution shows, the product $\sin \alpha_1 \sin \alpha_2 \sin \alpha_3$ is equal to $(r_1 r_2 r_3)/(R_1 R_2 R_3)$ where the R_i are the distances from P to the vertices of $\triangle ABC$ and the r_i are the distances to the sides. It is of interest that the maximum of $(r_1 r_2 r_3)/(R_1 R_2 R_3)$ is known to be $r/(4R)$, where r is the inradius and R the circumradius, in case P is constrained to the interior of $\triangle ABC$ (see D. S. Mitrinović, J. E. Pečarić and V. Volenec, *Recent Advances in Geometric Inequalities*, Kluwer, 1989). The proposers have shown that the maximum is $r_a/4R$ when there is no constraint on P , where r_a is the exradius opposite the largest angle A . Erdős has shown that $(3/2)R \leq \max\{r_a, r_b, r_c\}$ (see O. Bottema, et. al., *Geometric Inequalities*, Wolters-Noordhoff, 1989), giving another way to see that the maximum Miquel ratio is at least $3/8$.

No other correct solutions were received.

Surface Area of a Rectangular Parallelepiped

10256 [1992, 872]. *Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.*

Let $A_i, A'_i (i = 1, 2, 3, 4)$ be the vertices of a rectangular parallelepiped \mathcal{P} , with A'_i diametrically opposite to A_i . Let P be any interior point of \mathcal{P} . Prove that

$$S \leq 2(PA_1 \cdot PA'_1 + PA_2 \cdot PA'_2 + PA_3 \cdot PA'_3 + PA_4 \cdot PA'_4)$$

where S denotes the surface area of \mathcal{P} .

Solution by Robin J. Chapman, University of Exeter, U. K., Choose Cartesian coordinates with P at the origin, and suppose that the faces of \mathcal{P} lie in the planes $x = -a, x = a, y = -b, y = b, z = -c$ and $z = c$. The inequality becomes

$$\begin{aligned} & \sqrt{a^2 + b^2 + c^2} \sqrt{A^2 + B^2 + C^2} + \sqrt{a^2 + b^2 + C^2} \sqrt{A^2 + B^2 + c^2} \\ & + \sqrt{a^2 + B^2 + c^2} \sqrt{A^2 + b^2 + C^2} + \sqrt{A^2 + b^2 + c^2} \sqrt{a^2 + B^2 + C^2} \\ & \geq AB + Ab + aB + ab + CA + Ca + cA + ca + BC + Bc + bC + bc. \end{aligned}$$

Now by the Cauchy-Schwarz inequality

$$\sqrt{a^2 + b^2 + c^2} \sqrt{A^2 + B^2 + C^2} \geq aB + bC + cA,$$

and also

$$\sqrt{a^2 + b^2 + c^2} \sqrt{A^2 + B^2 + C^2} \geq Ab + Bc + Ca.$$

Hence

$$\sqrt{a^2 + b^2 + c^2} \sqrt{A^2 + B^2 + C^2} \geq (aB + bC + cA + Ab + Bc + Ca)/2$$

and by adding the similar inequalities obtained from the other terms on the left hand side of the main inequality, we get the main inequality.

Editorial comment. No other solver used coordinates based at P to simplify the formulas. Also, note that it is not necessary to require P to be an interior point. The interpretation as surface area is possible whenever $A + a$, $B + b$, and $C + c$ are all positive, and this choice can be made for any P . One reader noted that the result is false if one reads the terms $PA_i \cdot PA'_i$ as inner products of vectors. In this interpretation, the sum of inner products is negative whenever P is an interior point. We apologize for not noticing that this confusion was possible.

Solved also by D. Carlson, J. Fukuta (Japan), H. Kappus (Switzerland), I. Kastanas, R. A. Simon (Chile), and the proposer. One incorrect solution was received.

Familiar Inequalities in Disguise

10257 [1992, 872]. *Proposed by José A. Cavanati, Centro de Investigacion en Matematicas, A. C., Guanajuato, Mexico.*

Let a, b be positive real numbers, α a nonnegative real number, and m a positive integer with $\alpha \leq m$. Show that

$$0 \leq \frac{(a+b)^\alpha}{(a^m+b^m)^{\alpha/m}} - 1 \leq \frac{\alpha}{m} \left(\frac{(a+b)^m}{a^m+b^m} - 1 \right).$$

Solution by Eugene A. Herman, Grinnell College, Grinnell, IA. If we let $\beta = \alpha/m$, we can write the desired inequalities in the more revealing form

$$0 \leq \left(\frac{(a+b)^m}{a^m+b^m} \right)^\beta - 1 \leq \beta \left(\frac{(a+b)^m}{a^m+b^m} - 1 \right).$$

These are immediate consequences of the two elementary inequalities:

$$a^m + b^m \leq (a+b)^m \text{ whenever } m \geq 1, a, b > 0;$$

$$x^\beta - 1 \leq \beta(x - 1) \text{ whenever } x \geq 1, 0 \leq \beta \leq 1.$$

Note also that m need not be an integer; we require only that $m \geq 1$.

Solved by 47 readers and the proposer.

Collaborating editors: David F. Appleyard, Paul T. Bateman, Duane M. Broline, Barry W. Brunson, Frank S. Cater, Gulbank D. Chakerian, Underwood Dudley, Gerald A. Edgar, Michael A. Filaseta, Ira M. Gessel, Richard A. Gibbs, Jerrold R. Griggs, Douglas A. Hensley, John R. Isbell, Mourad E. H. Ismail, Murray Klamkin, Daniel J. Kleitman, Frederick W. Luttman, Frank B. Miles, Richard Pfiefer, Stephen L. Portnoy, J. O. Shallit, John Henry Steelman, Kenneth B. Stolarsky, David E. Tepper, Douglas B. Tyler, Daniel Ullman, and William E. Watkins.

REVIEWS

Edited by **Darrell Haile**

Indiana University, Bloomington IN 47405

The Lure of the Integers By Joe Roberts, Math. Assn. of America,
Washington D.C., 1992, v + 310, \$25

Reviewed by **Paul T. Bateman and Harold G. Diamond**

Mention an automobile model to a car enthusiast and you touch a memory—an uncle had that kind; another was an engineering marvel; a third was a commercial disaster. Discussion of food or wine will trigger a multitude of recollections in the gourmet. Say a number to the arithmetically inclined and you summon up thoughts on how it occurs in mathematics. For example, 1729 is the smallest number expressible as a sum of two cubes in two different ways, as was noted by Frenicle de Bessy and later by Ramanujan. Fascination with the properties of individual numbers may be as old as mathematics itself. Nicomachus gave the first four perfect numbers 6, 28, 496, 8128 around 100 A.D.; probably this list was known at least as far back as the time of Euclid [1].

More recently, this interest has led to the publication of several books on properties of individual numbers. One of the first, best known, and most accessible is Constance Reid's *From Zero to Infinity* [3], which focuses mainly on properties of the integers 0 through 9. *The Penguin Dictionary of Curious and Interesting Numbers* by David Wells [5] considers a much larger collection of numbers than does Reid, but rather more briefly. Closest in scope and spirit to Roberts' work is *Les Nombres Remarquables* by Francois le Lionnais [2]. Unlike Roberts, both Wells and le Lionnais discuss real numbers in general as well as integers.

The material in *The Lure of the Integers* ranges from the whimsical, e.g. Richard Guy's schoolgirl story (86)*, to the profound, e.g. Matijasevič's prime producing polynomial (5). Most of the topics in *Lure* come from the fields of recreational mathematics, number theory, and combinatorics. As a reflection of these interests, the names cited most frequently in the index are W. Sierpiński, P. Erdős, and D. H. Lehmer. Readers will of course find some sections more appealing than others; since there is essentially no connection between different topics, those which do not delight simply can be omitted.

What makes a number interesting? This is a question that we cannot answer, for interest is largely a matter of personal taste. It is unlikely that two people's lists of interesting numbers would be identical or that each number which appeared on two lists would be chosen for the same reason. The number 39 appears in

*A number in parentheses refers to a section of the book.

le Lionnais' collection as the first positive integer not possessing an interesting property (aside from *this* one). An inductive argument can be based on this remark to show that all positive integers are interesting.

Here are a few of the many items in *Lure* that we found attractive: (8) Schoenberg's theorem [4] that a regular simplex exists in \mathbb{Z}^n if and only if $n + 1$ is an odd square, a sum of two odd squares, or a multiple of 4; (163) the amazing irrational approximations to integers based on the fact that $\mathbb{Q}(\sqrt{-163})$ has unique factorization; and (561) Carmichael numbers, a sparse infinite set of composite numbers n having the property that $a^n - a$ is divisible by n for every integer a .

Exercising our right to disagree on what is interesting, we mention some topics which did not appeal to us: We were not captivated by digital properties of integers, such as the fact that 27 is the largest number equal to the sum of the digits of its cube (27); since the sum of the digits of n^k is less than $9(k \log_{10} n + 1)$, an infinite string of such assertions must hold. And our interest was not piqued by lists of exceptions to unsharp inequalities, such as the fact that 12 is the largest value for which the inequality $\tau(n) < n^{2/3}$ fails to hold (12), where $\tau(n)$ represents the number of positive integers dividing n . Indeed, since $\tau(n) = O(n^\delta)$, for any fixed $\delta > 0$, an infinite number of results of this type can be given.

It is incumbent on reviewers to point out at least a few items that ought to be changed. Here is our list: The reference to the first reviewer in Section 18.1 omitted his coauthor, P. Erdős. In the formula for $B(x)$ in Section 12.2, $\sqrt{\log x}$ should appear in place of the first $\log x$. Following this correction, the sentence "He showed that $NH(x)$ has the same form except that the first $\log x$ is changed to $\sqrt{\log x}$ " should be amended to end with the word "form." In Section 6.2 reference is made to a solution of *Monthly* Advanced Problem 5735. We note that this solution was subsequently found to be incorrect, and an explanation and correction were given in *This Monthly*, Vol. 97, page 937, 1990.

Roberts has a concise writing style. He provides the reader with statements, comments, and references but very few proofs, hints, or even definitions. Readers who are unacquainted with a topic will generally have to look elsewhere for an introduction. It is worth noting, as the author himself does, that the level of difficulty varies significantly among the topics and it may not be easy to decide whether an assertion is closer in difficulty to the triangle inequality or the Riemann hypothesis.

The strong point of *Lure* is the wealth of material it contains. The interested reader can test this assertion by the following experiment: write down several propositions each featuring a specific integer and see how many of them occur in the book. We found that this volume covers a significant proportion of the results we were familiar with. Also, the bibliographic documentation is very extensive, giving the book considerable value as a reference resource. *The Lure of the Integers* is a welcome contribution to the field of number-related literature.

REFERENCES

1. T. L. Heath, *The Thirteen Books of Euclid's Elements*, 2nd ed., Cambridge, 1926, reprinted by Dover, N.Y., 1956. Vol. 2, 425.
2. F. le Lionnais in collaboration with J. Brette, *Les Nombres Remarquables*, Actualités sci. et indust., 1407, Hermann, Paris, 1983.
3. C. Reid, *From Zero to Infinity*, 4th ed., MAA, Washington, D.C., 1992.

4. I. J. Schoenberg, Regular simplices and quadratic forms, J. London Math. Soc. 12 (1937), 48–55.
5. D. Wells, *The Penguin Dictionary of Curious and Interesting Numbers*, Penguin Books, London, 1986.

Department of Mathematics
University of Illinois
1409 W. Green St.
Urbana, IL 61801
diamond@math.uiuc.edu

Excursions in Calculus: An Interplay of the Continuous and the Discrete. By Robert M. Young, Mathematical Association of America, 1992, ix + 417, \$36.00.

Reviewed by Anita E. Solow

Calculus, Calculus, Calculus! Sometimes it seems that all I hear about these days is calculus. Over the past six years, there have been grants to reform calculus, books published to help us teach the course, and conferences to spread the word to both the initiated and the uninitiated. I have spent a good part of the past six years of my life concerned with the issues of calculus reform. So, when I was asked to review *Excursions in Calculus* by Robert Young, I was all set to write the definitive essay on calculus and the way it should be taught. After all, the title seemed to imply that this book has something to do with calculus. The preface strengthened that belief: “*Excursions in Calculus* is one possible supplement to a more traditional calculus course.” However, like any conscientious reviewer, my first step was to actually read the book. And when I did that, I realized that this book is not a calculus supplement at all.

Excursions in Calculus is a beautifully written, fascinating book. The book consists of six essays on the topics of Mathematical Induction, the Binomial Theorem, Fibonacci numbers, Averages, Approximations, and Infinite Series. Interwoven through these chapters are elementary number theory, combinatorics, probability theory, calculus, and geometry. Each chapter includes the history and development of these ideas, along with the connections among them. And each chapter is a gem. The writing is clear and compelling. The exercises at the end of each section are excellent. These include difficult, and even unsolved, problems that are fun to tackle and that develop many of the most interesting and historically famous results. For example, you will find developed in the problems the “ $3x + 1$ Problem,” Derangements, Public Key Cryptography, Stirling Numbers, Bell Numbers, Partitions, Buffon’s Needle Problem, Catalan Numbers, Stirling’s Formula, Bézier splines, Egyptian fractions, the Cantor Function, and the Sierpiński Triangle.

However, as wonderful as this book is, there are several reasons why it is not a supplement for calculus. First, I know of precious few first-year math students who could read this book. It is not that the book requires many prerequisites. After all, it assumes no mathematics above elementary calculus. However, the level of the exposition is too sophisticated for beginning mathematics students. The second problem is that the topics of this book are, for the most part, not really suitable for

inclusion in a calculus course, even if it were used only by the instructor to enrich the lectures. In the Preface, Paul Halmos is quoted as saying “there is no such subject as calculus; it is not a subject because it is many subjects.” I agree that calculus contains many ideas that students have seen previously, and many that they will see again at a higher level. However, I do believe that there is a core of material that makes up the subject of calculus, and I do not think it makes sense to change our calculus courses by throwing in lots of other mathematical topics.

The more I thought about it, the more I decided that this book is mistitled. One of my colleagues even suggested that a more appropriate title might be *Excursions out of Calculus*, since this book ties the continuous ideas of elementary calculus with the discrete ideas of number theory and combinatorics. Another, more appropriate, title might be *Excursions into Mathematics*, because of the wide variety of mathematical ideas that are explored. However, the latter is the title of a not dissimilar 1969 book by Anatole Beck, Michael Bleicher, and Donald Crowe [1]. It may be instructive to briefly compare these two books. On the surface there are many similarities. Both books consist of six essays on mathematical topics, and there is a nontrivial intersection of topics, with Beck containing more emphasis on geometry. Since Young’s book presumes a knowledge of calculus and the other does not, Young is able to delve into more advanced topics. The biggest difference in my mind is the structure of the chapters. *Excursions into Mathematics* contains six independent chapters. One of the strengths of Young’s *Excursions in Calculus* is the interdependence of the mathematics in the chapters. For example, ideas about prime numbers occur throughout the book, and are not neatly segregated into one area.

After I finished reading this book, I began to wonder how it could be used by students. It was clear to me that it should be read by them. But, when I looked at this book and looked at our mathematics curriculum, I could not get the two to mesh. *Excursions in Calculus* is full of beautiful, elegant mathematical ideas that most of us would agree should be part of the common culture of math students. I think that we would agree that all of our students should know some number theory and should have heard of the famous unsolved problems of the field, such as the Goldbach Conjecture and Fermat’s Last Theorem. Our students should understand Euler’s original argument that $\sum 1/n^2 = \pi^2/6$ and what the shortcomings of that argument were. They should understand the idea of approximation, whether it be approximating the value of π , or the idea of approximating a function by a polynomial, or the idea that $n \log n$ approximates the n^{th} prime number. They should be familiar with the search for prime numbers and should see the Bernoulli numbers in the context of summing the n^{th} powers of the integers.

Unfortunately, many of these ideas do not appear in the formal mathematics courses that we teach. Our students will see some of these ideas more by chance than by design. I believe that this is largely due to the separation of our discipline into distinct subdisciplines. So, when we teach combinatorics, we do not discuss number theory, and when we look at the idea of approximation in calculus, we do not look at approximations in number theory or numerical analysis. We all understand why our curriculum is structured the way it is, but it is unfortunate that the effect of this separation is to minimize the probability that the student will see the connections among these ideas.

Excursions in Calculus could be used to counter this deficiency in our curriculum. I could envision this book as the basis for a mathematical seminar for all math majors. Depending on the sophistication and talent of the students, the book could

easily be supplemented by more advanced writings on particular topics. At the joint AMS–MAA Mathematics Meetings in San Antonio (1993), there were well attended sessions on “capstone” courses for senior math majors. I would like to suggest this book as the foundation for such a course. Because of the variety of topics in the book, and because the purpose of the book is to put ideas from various mathematics courses together, it would serve the purpose of providing the structure upon which to hang a fascinating senior mathematical experience.

There are, of course, other ways to use this book; I encourage you to find one. All of our mathematics majors, not to mention all of us, would benefit from reading this book.

REFERENCES

1. Anatole Beck, Michael N. Bleicher, Donald W. Crowe, *Excursions into Mathematics*, Worth Publishers, Inc., New York, 1969.

Department of Mathematics & Computer Science
Grinnell College
Grinnell, IA 50112
solow@ac.grin.edu

Like Poetry, Mathematics is Beautiful

Timidly I ask
each one I meet if they
find mathematics beautiful
or useful, and each one dares to say,
“Useful, of course. I use it every day.”
And if I seem to want a proof,
they all go on to tell
that daily they subtract and add
to keep a checkbook; sometimes also
they multiply to find how many squares
they need to tile the kitchen floor.

Mathematics is not only plus
and minus, not just counting one,
two, three. There are rules to bend
defiantly, so parallels
will meet before infinity. Look
at the magic of unending terms
that converge to a finite sum:
start with one-half plus half of one-half
plus half of the last again and again.
Though we go on forever, we never
Pass one. Do you find me difficult? Oh, dear!

Suppose, instead, I ask
if poetry is beautiful
or useful. Will each person say,
“Useful, of course. I use it every day.”
And if I seem to want a proof,
they will go on to say that they
use rhymes to call to mind the days
of a month—like “Thirty hath
September”—and to remember
how to spell words with “i” and “e”.

I have a faint, enduring hope
that someday folks will see
mathematics to be
as lovely
as poetry.

JoAnne Growney
Department of Mathematics and Computer Science
Bloomsburg University / Bloomsburg, PA 17815

TELEGRAPHIC REVIEWS

Edited by **Arnold Ostebee and Paul Zorn**

with the assistance of the Mathematics Departments of
Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

<i>T</i> : Textbook	<i>P</i> : Professional Reading	1-4: Semester
<i>C</i> : Computer Software	<i>L</i> : Undergraduate Library	** : Special Emphasis
<i>S</i> : Supplementary Reading	13: Grade Level	?? : Questionable

Readers are advised that price information is subject to change. Selected books and software packages receive a second, more extensive review in the *Monthly*.

Books and software submitted for review should be sent to *Book Reviews Editor*, *American Mathematical Monthly*, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098.

General, S*(14-18), L.** *Essays in Humanistic Mathematics*. Ed: Alvin M. White. MAA Notes No. 32. MAA, 1993, xii + 212 pp, \$24 (P). [ISBN 0-88385-089-3] 22 brief, readable, sometimes provocative, sometimes elegant essays on broadly cultural aspects of mathematics—doing, learning, teaching. Mathematics is more than formal systems, say these authors: like the other humanities, mathematics reflects social, cultural, intellectual, and psychological contexts. A valuable, unusual resource. PZ

General, S(13-14), L. *Mathsemantics: Making Numbers Talk Sense*. Edward MacNeal. Penguin USA, 1994, 294 pp, \$22.95 (P). [ISBN 0-670-85390-9] Numeracy, says the author, requires attaching context-appropriate meanings ("mathsemantics") to otherwise formal numbers and operations. The case is made, and remedies suggested, in an anecdotal, often amusing, sometimes rambling style, built on an extended analysis of 200 job applicants' performance on a quantitative reasoning test. PZ

Reference, L.** *Table of Integrals, Series, and Products, Fifth Edition*. I.S. Gradshteyn, I.M. Ryzhik. Ed: Alan Jeffrey. Transl: Scripta Technica, Inc. Academic Pr, 1994, xlvii + 1204 pp, \$54.95. [ISBN 0-12-294755-X] Corrected and expanded. Many new entries and sections. (1980 edition, TR, December 1980.) AO

Mathematics Appreciation, T(13: 1). *Mathematical Thinking in a Quantitative World*. Linda R. Sons, Peter J. Nicholls. Kendall/Hunt, 1992, 289 pp, \$34.95 (P). [ISBN 0-8403-7924-2] Text for quantitative reasoning course: cov-

ers statistics, logical statements and arguments, graphical solution of systems of equations and inequalities, extrema of linear and quadratic functions, average rate of change, error analysis, business applications. Extensive exercise sets; good, nonroutine problems. Assumes elementary algebra, geometry. KES

Mathematics Appreciation, S(13-14), L. *Invitation to Mathematics*. Konrad Jacobs. Princeton Univ Pr, 1992, xi + 247 pp, \$29.95 (P); \$60. [ISBN 0-691-02528-2; 0-691-08567-6] Wide-ranging introduction to mathematical thinking. Topics include possible and impossible constructions, symmetry groups, Marriage Theorem, network flows, games, knots, fixed point theorems, dynamical systems. No exercises. Translation of 1987 German version. KES

Recreational Mathematics, S(13), L. *A Mathematical Pandora's Box*. Brian Bolt. Cambridge Univ Pr, 1993, 126 pp, \$15.95 (P). [ISBN 0-521-44619-8] 142 puzzles, games, tricks, and recreations, with solutions, at same level as in author's three previous puzzle books (*Amazing Mathematical Amusement Arcade*; *The Mathematical Funfair*, TR, March 1991; and *Mathematical Cavalcade*). LCL

Recreational Mathematics, P. *Mathematical Magic*. William Simon. Dover, 1993, 187 pp, \$5.95 (P). [ISBN 0-486-27593-0] Clever tricks that require little mathematics beyond arithmetic. An entertaining collection, well worth the price. Topics: magic with numbers, magic of shape, calendar magic, magic squares, and magic with playing cards. MK

Elementary, S(13). *Mathématiques générales: Problèmes résolus.* Jacques Bair, Geneviève Hamende. De Boeck Université (avenue Louise 203, B-1050, Bruxelles), 1992, iii + 236 pp, 145 FF (P). [ISBN 2-8041-1603-4] Problems and solutions in elementary mathematics: set theory, combinatorics, maxima and minima, exponential and logarithmic functions, etc. Accompanies authors' *Mathématiques générales* (2ème édition): 320 exercices, 380 problèmes avec leurs solutions, 160 tableaux. LC

Precalculus, T(13: 1, 2). *Mathematics for Calculus, Second Edition.* James Stewart, Lothar Redlin, Saleem Watson. Brooks/Cole, 1993, xviii + 773 pp, \$51.50. [ISBN 0-534-20250-0] Many new examples, exercises that use technology. Chapters end with problem-solving highlights. Margins include biographies, applications. (First Edition, TR, October 1989.) TH

Precalculus, T(13: 1, 2). *College Algebra and Trigonometry, Third Edition.* Bernard Kolman, Michael L. Levitan, Arnold Shapiro. Saunders College, 1993, xx + 768 pp, \$50.75. [ISBN 0-03-046933-3] New to this edition: scientific notation, critical value method for solving inequalities, partial fractions, determinants, polynomial functions, new material on analytic geometry. Over 350 graphing calculator problems. Features progress checks, carefully worked examples, warnings on pitfalls. (Second Edition, TR, November 1987.) TH

Education, P. *Street Mathematics and School Mathematics.* Terezinha Nunes, Analucia Dias Schliemann, David William Carraher. Cambridge Univ Pr, 1993, viii + 170 pp, \$49.95; \$16.95 (P). [ISBN 0-521-38116-9; 0-521-38813-9] Comparison of mathematics use in and out of school, across wide range of ages, cultures, and occupations. Concludes that "street mathematics is not the learning of particular procedures repeated in an automatic unthinking way, but involves the development of mathematical concepts and processes." Primary lesson for "realistic mathematics" teaching: respect the social context of problems. MW

Education, T(16-18: 1), S. *Problem Posing: Reflections and Applications.* Eds: Stephen I. Brown, Marion I. Walter. Lawrence Erlbaum Assoc, 1993, xvii + 336 pp, \$24.95 (P). [ISBN 0-8058-1065-X] 30 essays reflect on problem-posing rationale and strategies, primarily in pre-college mathematics. Excellent text for pre-service or in-service teachers. MW

Education, P. *Japanese Grades 7-9 Mathematics.* Ed: Kunihiro Kodaira. Transl: Hiromi Nagata, George Fowler. UCSMP Textbook Transl. Univ of Chicago School Math Project,

1992. Grade 7, xii + 185 pp [ISBN 0-936745-53-3]; Grade 8, xii + 205 pp [ISBN 0-936745-54-1]; Grade 9, xii + 199 pp. [ISBN 0-936745-55-X] Example-driven presentations of topics that U.S. students see in grades 7-11. Virtually all Japanese students complete grade 9; all are responsible for material in these texts. The lesson here is the depth to which all ninth graders are expected to learn mathematics. Don't look for innovative content or teaching ideas (manual calculation of square roots is a ninth grade Advanced Topic for Individual Study). MW

Education, T(13-14: 2). *Mathematics for Elementary Teachers: An Interactive Approach.* Thomas Sonabend. Saunders College, 1993, xviii + 917 pp, \$46.75. [ISBN 0-03-020709-6] Emphasis on discovery, discussion, applications, and BASIC programming. Each chapter ends with selected NCTM Curriculum Standards for students to connect to chapter material and to current elementary textbooks. Traditional table of contents. MW

Education, P. *How to Use Children's Literature to Teach Mathematics.* Rosamond Welchman-Tischler. NCTM, 1992, iv + 75 pp, \$8.50 (P). [ISBN 0-87353-349-6] Excellent resource for elementary teachers. Classroom activities, designs for worksheets and activity cards, and follow-up activities for twenty books. Sketches ideas for using additional books. KES

Education, P. *Rational Numbers: An Integration of Research.* Eds: Thomas P. Carpenter, Elizabeth Fennema, Thomas A. Romberg. Stud. in Math. Thinking & Learning. Lawrence Erlbaum Assoc, 1993, xi + 372 pp, \$79.95. [ISBN 0-8058-1135-4] 13 chapters report on research programs that address teaching, learning, curriculum, and assessment from an integrated perspective. Aims to "understand the effects of instruction" in rational numbers, not to "develop prescriptions for more effective instruction." Chapters on benefits of research linking teaching and learning are relevant throughout mathematics. MW

Education, P. *The Wonderful World of Mathematics: A Critically Annotated List of Children's Books in Mathematics.* Eds: Diane Thiessen, Margaret Matthias. NCTM, 1992, xi + 241 pp, \$17 (P). [ISBN 0-87353-353-4] Describes and rates usefulness of almost 500 trade books that emphasize mathematics concepts. Organized by concept. KES

History, S(13-17). *In the Wake of Chaos: Unpredictable Order in Dynamical Systems.* Stephen H. Kellert. Univ of Chicago Pr, 1993, xiv + 176 pp, \$19.95. [ISBN 0-226-42974-1] By a philosopher; starts with readable accounts

of both mathematics and history of dynamical systems. Later chapters on philosophical issues: "determinism" vs. "predictability;" social and cultural biases in research, etc. KS

Logic, S(15–16), L. *Popular Lectures on Mathematical Logic*. Hao Wang. Dover, 1993, vi + 281 pp, \$8.95 (P). [ISBN 0-486-67632-3] Reprint of 1981 edition (with new postscript). A wide-ranging treatment of logic and its connections to other areas of mathematics, computer science, and applications. Part technical, part philosophical. RM

Combinatorics, T(16–18: 2), S, P, L. *A Course in Combinatorics*. J.H. van Lint, R.M. Wilson. Cambridge Univ Pr, 1992, xii + 530 pp, \$80; \$29.95 (P). [ISBN 0-521-41057-6; 0-521-42260-4] Sophisticated introduction; abstract algebra recommended. Topics include Ramsey theory, extremal graphs, generating functions, codes and designs, association schemes, algebraic graph theory. AD

Combinatorics, P. *Oriented Matroids*. Anders Björner, et al. Ency. of Math. & Its Applic., V. 46. Cambridge Univ Pr, 1993, xii + 516 pp, \$89.95. [ISBN 0-521-41836-4] For experts and novices. Chapters treat axiomatics, face lattices, topological models, realizability, convex polytopes, linear programming. Complete in motivating idea of oriented matroids, in presenting topics, in exercises, and in bibliography. A thoroughly thorough book. SG

Number Theory, P. *Elementary Theory of L-functions and Eisenstein Series*. Haruzo Hida. London Math. Soc. Student Texts, V. 26. Cambridge Univ Pr, 1993, xi + 386 pp, \$69.95; \$22.95 (P). [ISBN 0-521-43411-4; 0-521-43569-2] A thorough introduction to p -adic modular forms and L -functions including Eichler-Shimura isomorphisms and functional equations for Hecke L -functions. SG

Number Theory, P. *Computational Algebraic Number Theory*. Michael E. Pohst. DMV Seminar, B. 21. Birkhäuser, 1993, 88 pp, \$26.50 (P). [ISBN 0-8176-2913-0] Describes algorithms developed recently to calculate a basis for the ring of integers in an algebraic number field, its unit group, and its ideal class group. SG

Linear Algebra, T(14: 1). *Linear Algebra for Mathematics, Science, and Engineering*. Edward M. Landesman, Magnus R. Hestenes. Prentice Hall, 1992, xiii + 551 pp. [ISBN 0-13-529561-0] Standard introductory text. Covers eigenvalues and eigenvectors before abstract vector spaces, linear transformations. Treats Gerschgorin intervals and disks, least squares, pseudoinverses, Rayleigh quotients, Gauss-Seidel methods. LC

Group Theory, P. *Geometric Group Theory, Volume 1*. Eds: Graham A. Niblo, Martin A. Roller. London Math. Soc. Lect. Note Ser., V. 181. Cambridge Univ Pr, 1993, 212 pp, \$34.95 (P). [ISBN 0-521-43529-3] Proceedings of 1991 Sussex University symposium.

Algebra, T(16–17: 2), S, L. *Abstract Algebra with Applications, in Two Volumes*. Karlheinz Spindler. Marcel Dekker, 1994, \$75 each, \$125 set. *Volume I: Vector Spaces and Groups*, xvii + 756 pp [ISBN 0-8247-9144-4]; *Volume II: Rings and Fields*, xv + 531 pp. [ISBN 0-8247-9159-2] A self-contained graduate text stressing clarity, not brevity. With historical motivation, concrete examples, informal comments. "Applications" of algebraic concepts are to geometry, topology, Markov chains, differential equations, combinatorics, number theory, and algebraic geometry. 58 chapters, each with many exercises. LCL

Algebra, T(14–16: 1, 2), L. *Abstract Algebra: A First Undergraduate Course, Fifth Edition*. Abraham P. Hillman, Gerald L. Alexanderson. PWS, 1994, xv + 480 pp. [ISBN 0-534-19128-2] Relatively minor changes in what has been (and will be) a widely used and highly regarded text. (*Fourth Edition*, TR, January 1989.) JS

Algebra, T(17–18: 2), L. *Algebra*. Mark Steinberger. PWS, 1994, xvi + 558 pp. [ISBN 0-534-93678-4] Elementary group theory; Sylow theory; categories; rings; field extensions; tensor products; linear algebra; Galois theory; semi-simple and hereditary rings; Dedekind domains. Assumes relatively little background. JS

Algebra, T(17), P*. *Noncommutative Algebra*. Benson Farb, R. Keith Dennis. Grad. Texts in Math., V. 144. Springer-Verlag, 1993, xiv + 223 pp, \$34. [ISBN 0-387-94057-X] A nice introduction to noncommutative algebra from a homological point of view. Core consists of Wedderburn structure theorems, Jacobson radical, double-centralizer theorems, and the Brauer group. Further topics include representation theory, primitive rings, global dimension. With a wealth of exercises. TH

Calculus, T(13: 1, 2). *Calculus, Part I & II, Revised Edition*. George W. Best, David A. Penner (Phillips Academy, Andover, MA 01810), 1992, \$45 set (P). *Part I*, 293 pp; *Part II*, 854 pp. Concepts precede techniques: introduces the definite integral and the derivative function before limits, differentiation techniques, or the fundamental theorem. Covers traditional single-variable calculus topics as well as parametric equations and vectors, differential equations, and sequences and series. AO

Calculus, S(13). *Exploring Precalculus and Calculus with the TI-81 Graphics Calculator.* George W. Best, David A. Penner. Venture Pub, 1992, 188 pp, \$25 (P). Exercises and projects on functions, derivatives, and integrals. Emphasizes graphical and numerical approaches. First chapter introduces calculator capabilities, including programming. AO

Calculus, T(13: 2). *Technical Calculus with Analytic Geometry, Third Edition.* Peter Kuhfittig. Brooks/Cole, 1994, xii + 525 pp, \$65. [ISBN 0-534-21852-0] Nonrigorous treatment stressing applications. Three chapters on differential equations, including Laplace transforms. Exploits graphing calculators. HD

Real Analysis, P. *Lectures on Ergodic Theory and Pesin Theory on Compact Manifolds.* Mark Pollicott. London Math. Soc. Lect. Note Ser., V. 180. Cambridge Univ Pr, 1993, ix + 162 pp, \$37.95 (P). [ISBN 0-521-43593-5] First half covers basic ergodic theory (recurrence, ergodic theorems, entropy); remainder treats Pesin theory, the study of non-uniformly hyperbolic diffeomorphisms. SB

Complex Analysis, T(18), S, P, L. *Invariant Distances and Metrics in Complex Analysis.* Marek Jarnicki, Peter Pflug. Expos. in Math., V. 9. Walter de Gruyter, 1993, xi + 408 pp, DM 178. [ISBN 3-11-013251-6] Theory of holomorphically constructible objects on domains in n -dimensional complex space: Carathéodory and Kobayashi pseudodistances and pseudometrics, complex Green's function, Bergman distance and metric, etc. Many examples, exercises, open problems, new results. KS

Complex Analysis, T(18), L? *Rational Iteration: Complex Analytical Dynamical Systems.* Stud. in Math., V. 16. Walter de Gruyter, 1993, ix + 189 pp, DM 108. [ISBN 3-11-013765-8] A comprehensive account of substantive work of Fatou and Julia—so often alluded to in superficial treatments rushing to the now famous pictures. Quasiconformal mappings are de-emphasized, except as needed to sketch the idea of Sullivan's theorem on no wandering domains. Assumes basic complex analysis. AWR

Differential Equations, S(18), P. *Treatise on Analysis, Volume VIII.* J. Dieudonné. Transl: Laura Fainsilber. Pure & Appl. Math., V. 10-VIII. Academic Pr, 1993, xi + 356 pp, \$110. [ISBN 0-12-215508-4] Chapter 23, Part II of Dieudonné's *magnum opus*, concentrating on boundary value problems. The style is well known, which is to say dense—not a gentle introduction for the uninitiated. AWR

Differential Equations, T*(18: 2), P. *Applications of Lie Groups to Differential Equations,*

Second Edition. Peter J. Olver. Grad. Texts in Math., V. 107. Springer-Verlag, 1993, xxviii + 513 pp, \$59. [ISBN 0-387-94007-3] The book to read to learn about symmetry groups of differential equations and their many applications (solving PDE's and ODE's explicitly, finding conservation laws, etc.). New edition adds brief discussion of pseudo-differential operators, and maintains a wealth of examples, frank and interesting historical notes, and a most readable expository style. JO

Differential Equations, T(14–15: 1). *Ordinary Differential Equations with Applications, Third Edition.* Bernard J. Rice, Jerry D. Strange. Brooks/Cole, 1994, ix + 533 pp, \$63. [ISBN 0-534-21318-9] Well-chosen examples, clear writing style, many applications. Stresses linear equations. Technology is used (and recommended) throughout. DS

Differential Equations, T(18: 1), P. *Delay Differential Equations With Applications in Population Dynamics.* Yang Kuang. Math. in Sci. & Eng., V. 191. Academic Pr, 1993, xii + 398 pp, \$59.95. [ISBN 0-12-427610-5] Research results in population dynamics using delay differential equations. Reviews basic theory, motivated by applications. DS

Differential Equations, S(14–15). *Differential Equations with Mathematica.* Martha L. Abell, James P. Braselton. Academic Pr, 1993, viii + 631 pp, \$44.95 (P). [ISBN 0-12-041538-0] Supplements a standard ODE course; demonstrates uses of Mathematica. Not a full-fledged text; no exercises, insufficient exposition. JO

Partial Differential Equations, P. *Degenerate Parabolic Equations.* Emmanuele DiBenedetto. Universitext. Springer-Verlag, 1993, xv + 387 pp, \$39 (P). [ISBN 0-387-94020-0] Overview of recent results. JO

Partial Differential Equations, T(15: 1). *Partial Differential Equations for Scientists and Engineers.* Stanley J. Farlow. Dover, 1993, ix + 414 pp, \$12.95 (P). [ISBN 0-486-67620-X] A standard introduction, 47 “semi-independent” lessons. Unusual topics include control theory, calculus of variations, Monte Carlo methods, and potential theory. Relatively few exercises. (1982 Wiley edition, TR, January 1983.) JO

Partial Differential Equations, T(18), P. *Partial Differential Equations IV: Microlocal Analysis and Hyperbolic Equations.* Eds: Yu. V. Egorov, M.A. Shubin. Ency. of Math. Sci., V. 33. Springer-Verlag, 1993, 241 pp, \$79. [ISBN 0-387-53363-X] Two independent themes from current research in PDE's: (1) survey of results in local analysis of cotan-

gent bundle space; (2) linear hyperbolic equations and systems. DS

Dynamical Systems, P. *Hyperbolicity and Sensitive Chaotic Dynamics at Homoclinic Bifurcations: Fractal Dimensions and Infinitely Many Attractors.* Jacob Palis, Floris Takens. Stud. in Adv. Math., V. 35. Cambridge Univ Pr, 1993, x + 234 pp, \$54.95. [ISBN 0-521-39064-8] A self-contained introduction to homoclinic bifurcation theory. Begins with detailed treatments of several classical examples. Stresses interplay between homoclinic tangencies and non-trivial basic sets. Includes a geometric proof of Newhouse's Theorem. MPR

Numerical Analysis, P. *Mathematical and Computational Techniques for Multilevel Adaptive Methods.* Ulrich Rüde. Frontiers in Appl. Math., V. 13. SIAM, 1993, xii + 140 pp, \$23 (P). [ISBN 0-89871-320-X] Basic theory of multilevel adaptive methods for solving PDE's numerically. Discusses data structures needed for such methods, how object-oriented programming implements these data structures. JO

Operator Theory, T(17), L. *Hilbert Space: Compact Operators and the Trace Theorem.* J.R. Retherford. London Math. Soc. Stud. Texts, V. 27. Cambridge Univ Pr, 1993, xii + 131 pp, \$44.95; \$19.95 (P). [ISBN 0-521-41884-4; 0-521-42933-1] A readable introduction to Hilbert space and Banach space techniques, moving toward the Lidskij trace theorem as culmination. Aims to be comprehensible to beginning graduate students, useful to advanced students; assumes advanced calculus and linear algebra. AWR

Functional Analysis, T(16-17: 1, 2), S, L. *Applied Algebra and Functional Analysis.* Anthony N. Michel, Charles J. Herget. Dover, 1993, x + 484 pp, \$10.95 (P). [ISBN 0-486-67598-X] Good text for engineering and science students. Covers algebra (including Jordan canonical form), metric spaces, normed spaces, inner-product spaces, and operator theory through the spectral theorem for compact normal operators. Applications to differential and integral equations, random variables, optimal control, minimization. HD

Functional Analysis, P. *Function Spaces, Differential Operators and Nonlinear Analysis.* Eds: Hans-Jürgen Schmeisser, Hans Triebel. Teubner-Texte zur Mathematik, Band 133. BG Teubner Leipzig, 1993, 308 pp, DM 54,80 (P). [ISBN 3-8154-2045-8] Proceedings of a 1992 conference in Friedrichroda, Germany.

Analysis, T*, P*, L. *Harmonic Function Theory.* Sheldon Axler, Paul Bourdon, Wade Ramey. Grad. Texts in Math., V. 137. Springer-

Verlag, 1992, xii + 231 pp, \$39.50. [ISBN 0-387-97875-5] Harmonic function theory in \mathbb{R}^n , starting with the basics. Topics include Bôcher's theorem, spherical harmonics, harmonic Hardy and Bergman spaces, Dirichlet problem. Assumes "solid foundation in real and complex analysis, and some basic results from functional analysis." Engaging style; plenty of motivation. Many good problems end each chapter. Well worth a look. BH

Analysis, P, L. *Inverse Problems in the Mathematical Sciences.* Charles W. Groetsch. Friedr Vieweg & Sohn, 1993, v + 152 pp, \$30. [ISBN 3-528-06545-1] "Broad-based introduction" to "some of the main lines of research in inverse and ill-posed problems." Numerous physical models, sketch of some operator theory background, survey of some numerical methods. Extensive annotated bibliography. BH

Analysis, T(17), S, L. *Fundamentals of Convex Analysis: Duality, Separation, Representation, and Resolution.* Michael J. Panik. Theory & Decision Lib., Ser. B, V. 24. Kluwer Academic, 1993, xxii + 294 pp, \$108. [ISBN 0-7923-2279-7] Considerable overlap in coverage and viewpoint with Rockafellar's *Convex Analysis* (TR, April 1970), but written for students of economics, management science, and engineering rather than mathematicians. Nicely provides the theory necessary for linear programming, duality, separation theorems, etc. Some proofs are omitted. AWR

Analysis, T(17-18: 1, 2), S, L. *Problems in Real and Complex Analysis.* Bernard R. Gelbaum. Prob. Books in Math. Springer-Verlag, 1992, ix + 488 pp, \$59. [ISBN 0-387-97766-X] A large, useful collection of problems, with full solutions, in real and complex analysis. Difficulty ranges from routine to quite challenging. Excellent for self-study, preparation for doctoral prelims, professional brushing-up. PZ

Analysis, P. *Lie Groups and Lie Algebras I: Foundations of Lie Theory, Lie Transformation Groups.* Ed: A.L. Onishchik. Ency. of Math. Sci., V. 20. Springer-Verlag, 1993, 235 pp, \$89. [ISBN 0-387-18697-2] Two major chapters: Foundations of Lie Theory (covers basics of Lie groups and Lie algebras and their relationship, the universal enveloping algebra, and generalizations of Lie groups) and Lie Transformation Groups (considers Lie group actions on manifolds, various homogeneous spaces, and actions of compact Lie groups). SG

Optimization, T(16-17: 1, 2), L*. *Network Flows: Theory, Algorithms, and Applications.* Ravindra K. Ahuja, Thomas L. Magnanti, James B. Orlin. Prentice Hall, 1993, xv

+ 846 pp. [ISBN 0-13-617549-X] A comprehensive, up-to-date introduction to shortest path, maximum flow, and minimum cost flow problems. Relatively few computational exercises; most exercises address theory or applications. Extensive reference notes. AO

Mathematical Modeling, T(16-17: 1). *Mathematical Modelling of Complex Mechanical Systems, Volume 1: Discrete Models.* K. Arczewski, J. Pietrucha, C.M. Leech. Math. & Its Applic. Ellis Horwood, 1993, 293 pp, \$42. [ISBN 0-13-563750-3] Unified treatment of models from mechanical engineering, thermodynamics, fluid mechanics, etc. DS

Control Theory, P. *Nonlinear Feedback Control Systems: An Operator Theory Approach.* Rui J.P. de Figueiredo, Guanrong Chen. Academic Pr, 1993, ix + 220 pp, \$59.95. [ISBN 0-12-208630-9] Presents one approach to unifying mathematical results of nonlinear control theory. Aims for a mathematical theory that will facilitate the formulation of "... nonlinear problems, the understanding of the underlying system properties, and the construction of the pertinent algorithms." JO

Probability, T(13: 1). *Introduction to Probability.* John E. Freund. Dover, 1993, viii + 247 pp, \$6.95 (P). [ISBN 0-486-67549-1] Elementary, easy-reading text covers combinatorics, conditional probability, expectation, probability distributions, the law of large numbers. Republication of a 1973 Dickenson text (TR, August-September 1973). RWJ

Probability, T(15-16: 1). *Introduction to Probability Models, Fifth Edition.* Sheldon M. Ross. Academic Pr, 1993, xi + 556 pp, \$59.95. [ISBN 0-12-598455-3] This edition includes examples for greedy algorithms, minimizing highway encounters, tracking AIDS, compound Poisson process applications, and theory of options pricing including arbitrage theorem and Black-Scholes option pricing formula. 120 new exercises. (Fourth Edition, TR, February 1990.) MK

Probability, T(16-17: 1, 2), L. *Probability.* Alan F. Karr. Texts in Stat. Springer-Verlag, 1993, xxi + 282 pp, \$39. [ISBN 0-387-94071-5] Probability, random variables, independence, expectation, convergence of sequences of random variables, characteristic functions, limit theorems, prediction and conditional expectation, and martingales. Preliminary section on random walks serves as introduction. Assumes undergraduate real analysis. RWJ

Stochastic Processes, S(18), P. *Continuous-Time Markov Chains: An Applications-Oriented Approach.* William J. Anderson.

Ser. in Stat. Springer-Verlag, 1991, xii + 355 pp, \$69. [ISBN 0-387-97369-9] Theory of continuous-time Markov chains via transition functions, approached through backward and forward equations and integral recursions. Assumes background in analysis, probability, Markov chains. No exercises. MK

Stochastic Processes, T(13: 1). *An Introduction to Probability and Stochastic Processes.* Marc A. Berger. Texts in Stat. Springer-Verlag, 1993, xii + 205 pp, \$39. [ISBN 0-387-97784-8] Elegant presentation from a non-measure-theoretic viewpoint. Terse exposition, discussion, proofs; excellent problem sets. Topics include Markov chains, passage phenomena, stationary distributions and steady states, Markov jump processes, ergodic theory with applications to fractals. MK

Stochastic Processes, T(18: 2). *Brownian Motion and Stochastic Calculus, Second Edition.* Ioannis Karatzas, Steve E. Shreve. Grad. Texts in Math., V. 113. Springer-Verlag, 1991, xxiii + 470 pp, \$39.50 (P). [ISBN 0-387-97655-8] Advanced treatment of the continuous-time Markov (and Martingale) property. Brownian motion is studied thoroughly at outset, leading to stochastic integration, Brownian motion and PDE's, stochastic DE's, and Levy's theory of Brownian local time. MK

Stochastic Processes, T(16-17: 2). *An Introduction to Stochastic Modeling, Revised Edition.* Howard M. Taylor, Samuel Karlin. Academic Pr, 1994, xi + 566 pp, \$59.95. [ISBN 0-12-684885-8] Revisions include many more exercises (ordered by difficulty), Wald's equation now appears in renewal theory chapter, examples of sum quota sampling. (First Edition, TR, May 1985.) KB

Stochastic Processes, P. *Poisson Processes.* J.F.C. Kingman. Oxford Stud. in Prob., V. 3. Clarendon Pr, 1993, viii + 104 pp, \$39.95. [ISBN 0-19-853693-3] Surveys Poisson processes in one or more dimensions. Several applications to ecology. Avoids measure theory, but contains ample mathematical detail. Topics include marked Poisson processes, Cox processes, stochastic geometry, Poisson-Dirichlet distribution. No exercises. MK

Elementary Statistics, T(13: 1). *A First Course in Statistics, Fourth Edition.* James T. McClave, Frank H. Dietrich II. Dellen, 1992, xv + 583 pp. [ISBN 0-02-378561-6] Deletes "optional" topics from previous edition (TR, October 1983) to shorten text. Comparisons of means and proportions are treated in separate chapters. Numerous examples from current events and scientific research areas. Includes a

substantial demographic data set, many exercises. Emphasizes mechanics. MK

Elementary Statistics, T(13: 1). *Statistical Methods*. Rudolf J. Freund, William J. Wilson. Academic Pr, 1993, xix + 644 pp, \$52.50. [ISBN 0-12-267470-7] Non-calculus-based introduction to statistics. Emphasizes drawing conclusions from analyses, not mechanics. Many exercises. Covers topics through multiple linear regression, factorial experiments, weighted least-squares, logistic regression, and log-linear models. MK

Elementary Statistics, T(13: 1, 2), S, C. *Hyperstat: Macintosh Hypermedia for Analyzing Data and Learning Statistics*. David M. Lane. Academic Pr, 1993, x + 128 pp, \$59.95 (P), with disks. [ISBN 0-12-436130-7] Combines a statistics text, a data analysis program, and interactive exercises. Hypertext links related concepts. Main text is on disk; accompanying manual explains general use, specific statistical procedures, explorations options. Assumes Macintosh basics, but no prior knowledge of statistics. KB

Elementary Statistics, T(14: 1), L?? *Statistics in Theory and Practice*. Robert Lupton. Princeton Univ Pr, 1993, x + 188 pp, \$24.95. [ISBN 0-691-07429-1] A text for math and science majors. Assumes knowledge of calculus ideas and techniques. "There is very little real data in this book, and very few exercises of the sort that ask you to apply the techniques presented to real problems." MLR

Mathematical Statistics, T(14-15). *Foundations of Probability and Statistics*. William C. Rinaman. Saunders College, 1993, xii + 773 pp, \$58.75. [ISBN 0-03-071806-6] Stresses theory; few "real" data sets or case studies. Traditional contents. Examples, exercises are many and useful. MK

Mathematical Statistics, S(15-16). *Solutions in Statistics and Probability (Second Edition)*. Edward J. Dudewicz. Amer. Ser. in Math. & Management Sci., V. 3. American Sciences Pr, 1993, iv + 318 pp, \$98.75 (P). [ISBN 0-935950-35-4] Detailed solutions to problems in *Introduction to Statistics and Probability* by Dudewicz, and *Modern Mathematical Statistics* by Dudewicz and Mishra. (First Edition, TR, April 1982.) RWJ

Mathematical Statistics, T(14). *Probability and Statistics for Engineers and Scientists, Fifth Edition*. Ronald E. Walpole, Raymond H. Myers. Macmillan, 1993, xiv + 766 pp. [ISBN 0-02-424201-2] New edition emphasizes data analysis more than before (*Fourth Edition*, TR, December 1985). Includes new exercises em-

phasizing "real-life" science and engineering applications, and complete case studies with computer printout (MINITAB), graphics, and diagnostic measures. Section on Taguchi's robust parameter design. MK

Statistical Methods, T(17). *Time Series: Forecasting, Simulation, Applications*. Gareth Janacek, Louise Swift. Math. & Its Applic. Ellis Horwood (US Distr: Simon & Schuster), 1993, 331 pp, \$74.95. [ISBN 0-13-918459-7] Introduction to time series, theory and applications. State space models, Kalman filter are introduced early, with emphasis on exact maximum-likelihood estimation. These ideas are then recast for ARMA model development and for structural models; also treats frequency domain ideas. Special topics: multiple series, missing data, long-memory processes, etc. MK

Statistical Methods, T(16-17: 1), P, L. *Statistical Design and Analysis for Intercropping Experiments, Volume I: Two Crops*. Walter T. Federer. Ser. in Stat. Springer-Verlag, 1993, xx + 298 pp, \$59. [ISBN 0-387-97923-9] Surveys statistical design and analysis of intercropping experiments. Applies statistical models to relative indices, yield-density relations, biological competition, and spatial-intimacy arrangements. Includes several nice data sets. RWJ

Statistical Methods, P. *Bayesian Statistics 4*. Eds: J.M. Bernardo, et al. Clarendon Pr, 1992, xiii + 859 pp, \$95. [ISBN 0-19-852266-5] Proceedings of the Fourth Valencia International Meeting on Bayesian Statistics, April 1991. MK

Statistics, P. *Nonlinear Statistical Models*. Andrej Pázman. Math. & Its Applic., V. 254. Kluwer Academic, 1993, ix + 259 pp, \$119. [ISBN 0-7923-2247-9] Treats L_2 estimators in (multivariate) nonlinear regression models, their properties (using differential-geometrical methods), their computation, and local and global approximations of their probability densities. Introductory chapter on linear regression, concluding chapter on nonlinear exponential families. RWJ

Algorithms, P. *Efficient Algorithms for Listing Combinatorial Structures*. Leslie Ann Goldberg. Cambridge Univ Pr, 1993, xv + 160 pp, \$44.95. [ISBN 0-521-45021-7] Award-winning dissertation. LC

Computer Organization, P. *High Performance Computing*. Kevin Dowd. O'Reilly & Assoc, 1993, xxv + 371 pp, \$25.95 (P). [ISBN 1-56592-032-5] For "people who are interested in computer performance or who need to understand it for their job." Treats RISC architectures, modern memory systems, optimizing

compilers, parallel processing, bench-marking, porting code. A brisk, practical overview. JO
Computer Systems, S(16-17), P, C. *Introduction to Computer Performance Analysis with Mathematica*. Arnold O. Allen. Comp. Sci. & Scient. Computing. Academic Pr, 1994, xx + 356 pp, \$49.95 with disk. [ISBN 0-12-051070-7] Tools for modeling, simulation, and performance analysis of computer systems; all implemented in Mathematica, with code on accompanying disk. Suitable for self-study. RM

Computer Systems, P. *ORACLE Performance Tuning*. Peter Corrigan, Mark Gurry. O'Reilly & Assoc, 1993, xxxv + 603 pp, \$34.95 (P). [ISBN 1-56592-048-1]

Computer Graphics, P. *Quick Reference to Computer Graphics Terms*. Roger T. Stevens. Academic Pr, 1993, vi + 237 pp, \$29.95 (P). [ISBN 0-12-668310-7] A dictionary of graphics terms aimed at readers who, like the author, read graphics literature and find new terms and acronyms around every corner. A good idea. JO

Artificial Intelligence, T*(15-18: 1, 2), L. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving, Second Edition*. George F. Luger, William S. Stubblefield. Benjamin/Cummings, 1993, xxiv + 740 pp. [ISBN 0-8053-4780-1] Attempts to unify artificial intelligence theory and problem solving techniques. Stresses representational formalisms, search techniques; artificial intelligence within scientific tradition. New to this edition: machine learning, object-oriented design with the Common Lisp Object System. RJA

Computer Science, T(16-18: 1, 2), P, L. *Action Semantics*. Peter D. Mosses. Tracts in Theor. Comp. Sci., V. 26. Cambridge Univ Pr, 1992, xx + 372 pp, \$49.95. [ISBN 0-521-40347-2] Action semantics concepts and formalism, action notation, and examples of action semantic descriptions. Relates action semantics to other frameworks, briefly sketches its development. Appendices cover formal details, serve as complete, self-contained reference manual. RWJ

Computer Science, T(17-18: 1). *Algorithmic Algebra*. Bhubaneswar Mishra. Texts & Mono. in Comp. Sci. Springer-Verlag, 1993, xii + 416 pp, \$39.95. [ISBN 0-387-94090-1] Effective algorithms for computing various algebraic structures—Gröbner bases, characteristic sets, resultants, and semi-algebraic sets. For anyone who wants to understand the algorithmic underpinnings of computer algebra systems. DS

Applications (Engineering), C. *HP 48SX Engineering Mathematics Library*. John F. Hol-

land. Academic Pr, 1992, xxiv + 632 pp, \$139.95. [ISBN 0-12-352380-X] Book documents an immense library (over 700 programs, supplied on an HP memory card, included) of HP 48SX-based mathematical functions, programs, and applications, especially useful in engineering mathematics. Includes utilities for complex analysis, special functions, linear algebra, statistics, signal processing, differential equations, engineering applications, plotting of all sorts, etc. A reference manual both to software and (via appendices) to considerable amount of underlying mathematics. PZ

Applications (Fluid Dynamics), T(15-16: 1). *A Mathematical Introduction to Fluid Mechanics, Third Edition*. Alexandre J. Chorin, Jerrold E. Marsden. Texts in Appl. Math., V. 4. Springer-Verlag, 1993, xi + 169 pp, \$34. [ISBN 0-387-97918-2] Informal mathematical introduction to theory of fluid mechanics. Concepts are carefully motivated and developed, sometimes informally rigorous. Many examples, figures. (Second Edition, TR, February 1991.) DS

Applications (Physical Science), T(17: 2), P. *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*. Donald B. Percival, Andrew T. Walden. Cambridge Univ Pr, 1993, xxvii + 583 pp, \$89.95; \$39.95 (P). [ISBN 0-521-35532-X; 0-521-43541-2] Introduction to univariate spectral analysis. Covers traditional topics and recent advances. Numerous, useful exercises. Clear exposition; many examples motivate and implement the theory. MK

Applications (Physical Science), P. *Statistics and Physical Oceanography*. National Research Council. National Academy Pr, 1993, x + 62 pp, (P). Monograph on research problems in statistics and applied probability that arise in oceanography. Topics include statistical issues in multiple-scale variability of oceanographic fields; Lagrangian and Euler data and models; interpolation, non-linear smoothing, filtering, and prediction; model comparison; non-Gaussian random fields. MK

Reviewers

RJA: Richard J. Allen, St. Olaf; KB: Karla Ballman, Macalester; LC: Laura Chihara, St. Olaf; AD: Amy Davidow, Macalester; HD: Hung Dinh, Macalester; SG: Steven Galovich, Carleton; TH: Tom Halverson, Macalester; BH: Bruce Hanson, St. Olaf; RWJ: Roger W. Johnson, Carleton; MK: Michael Kahn, St. Olaf; LCL: Loren C. Larson, St. Olaf; RM: Richard Molnar, Macalester; JO: Jeff Ondich, Carleton; AO: Arnold Ostebee, St. Olaf; MLR: Margaret L. Reese, St. Olaf; MPR: Matthew P. Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; KS: Karen Saxe, Macalester; JS: John Schue, Macalester; DS: Dan Schwalbe, Macalester; KES: Kay E. Smith, St. Olaf; MW: Martha Wallace, St. Olaf; PZ: Paul Zorn, St. Olaf.

**No matter how you
express it, it still means
DERIVE® is half price.**

$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2}$
 $\lim_{x \rightarrow 0} \frac{x}{\sin(2x)}$
 $\frac{1}{2}$

50%

$\sum_{n=1}^{\infty} \frac{1}{2^{n+1}}$

0.5

$\int_0^1 x \, dx$

DERIVE →

The *DERIVE A Mathematical Assistant* program lets you express yourself symbolically, numerically and graphically, from algebra through calculus, with vectors and matrices too—all displayed with accepted math notation, or 2D and 3D plotting. *DERIVE* is also easy to use and easy to read, thanks to a friendly, menu-driven interface and split or

overlay windows that can display both algebra and plotting simultaneously. Better still, *DERIVE* has been praised for the accuracy and exactness of its solutions. But, best of all the suggested retail price is now only \$125. Which means *DERIVE* is now half price, no matter how you express it.

System requirements

DERIVE: MS-DOS 2.1 or later, 512K RAM, and one 3½" disk drive. Suggested retail price now **\$125 (Half off!)**.

DERIVE ROM card: Hewlett Packard 95LX & 100LX Palmtop, or other PC compatible ROM card computer. Suggested retail price now **\$125!**

DERIVE XM (eXtended Memory): 386 or 486 PC compatible with at least 2MB of *extended* memory. Suggested list price now \$250!

DERIVE is a registered trademark of Soft Warehouse, Inc.



Soft Warehouse
HONOLULU • HAWAII

Soft Warehouse, Inc. • 3660 Waiālae Ave.
Ste. 304 • Honolulu, HI, USA 96816-3236
Ph: (808) 734-5801 • Fax: (808) 735-1105

NEW IN THE SPECTRUM SERIES

The Words of Mathematics

An Etymological Dictionary of Mathematical Terms Used in English

Steven Schwartzman

Heteroscedastic

Amphicheiral

Centroid

Clothoid

Eigenvalue

The Words of Mathematics explains the origins of over 1500 mathematical terms used in English. While other dictionaries of mathematics define technical terms, this book concentrates on where those terms came from and what their literal meanings are. The words included here range from simple to advanced. Elementary school teachers may be surprised to learn that *inch* and *ounce* are really the same word, that *eleven* means literally "one left over," that a *thousand* is a "swollen hundred," and that the original meaning of *times* was "divide." High school teachers will find out that *asymptote* means "not falling together," that an *area* used to be a threshing floor, and that *focus* is a Latin word meaning "fireplace." College teachers who want to explain *heteroscedastic*, *amphicheiral*, and *eigenvalue* to their students will find the origins of those words in this book.

This dictionary is easy to use. Although some of the entries are highly technical, the book explains them in plain English. The introduction gives an overview of how the ancient language known as Indo-European developed into Latin, Greek, French, and English, the languages from which most of our mathematical vocabulary has been derived. Another section discusses the many ways

in which mathematicians have borrowed and created their specialized vocabulary over the centuries. A glossary explains historical and linguistic terms used throughout the book.

As in any dictionary, the entries themselves are arranged alphabetically. The words are drawn from arithmetic, algebra, geometry, trigonometry, calculus, number theory, topology, statistics, graph theory, logic, recreational mathematics, and other areas. Over 200 illustrations accompany the dictionary entries, especially some of the less familiar ones. Connections to related nonmathematical English words are often pointed out. Key numbers attached to many entries lead interested readers to an appendix which groups mathematical terms that come from a common source.

This dictionary is an indispensable reference for every library that serves teachers and students of mathematics. It is a natural source of information for courses in the history of mathematics and for mathematics courses intended for liberal arts students.

262 pp., Paperbound, 1994

ISBN 0-88385-511-9

List: \$27.00 MAA Member: \$21.00

Catalog Number WORDS

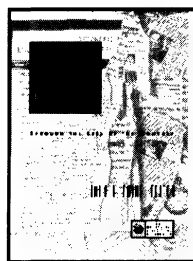
	Qty.	Catalog Number	Price

			Total \$
Name	-----		
Address	-----		
City	-----		
State	Zip Code	-----	
		Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD	
		Credit Card No.	-----
		Signature	-----
		Exp. Date	-----

NEW NEW NEW

A Century of Mathematics Through the Eyes of the Monthly

John Ewing, Editor



The Monthly is 100 years old!

*We are celebrating this special birthday with the publication of
A Century of Mathematics: Through the Eyes of the Monthly.*

This is the story of American mathematics during the past century. It contains articles and excerpts from a century of the **Monthly**, giving the reader an opportunity to skim all one hundred volumes without actually opening them. It samples mathematics year by year and decade by decade. Along the way, readers can glimpse the mathematical community at the turn of the century, and the divisions between the mathematical communities of teachers and researchers. They read about the struggle to prevent colleges from eliminating mathematics requirements in the 1920s, the controversy about Einstein and relativity, the debates about formalism in logic, the immigration of mathematicians from Europe, and the frantic effort to organize as the war began. At the end of the war, they hear about new divisions between pure and applied mathematics, heroic efforts to deal with large numbers of new students in the universities, and the rise of federal funding for mathematics. In more recent times, they see the advent of computers and computer science, the problems faced by women and minorities, and some of the triumphs of modern research.

This is a book about mathematics—about teaching and research, applied and pure, elite universities and community colleges. Browsing through its pages, readers see what has changed (the kinds of mathematics in fashion, for example) and what has stayed the same (our concern about teaching and our complaints about the deplorable state of our students).

This is a book about history—a sampling of history, meant to be savored rather than studied. For one hundred years, the **Monthly** has contained articles by some of the greatest mathematicians in the world, as well as articles by students and faculty from small midwestern colleges where those great names were barely known. This book gives a glimpse of both worlds. It tells a story rather than the details of history.

This is the story of a century of mathematics in America.

335 pp., Hardbound, 1994

ISBN 0-88385-457-0

List: \$ 39.50 MAA Member: \$32.00

Catalog Number CENTMA

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty.

Catalog Number

Price

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

Knot Theory

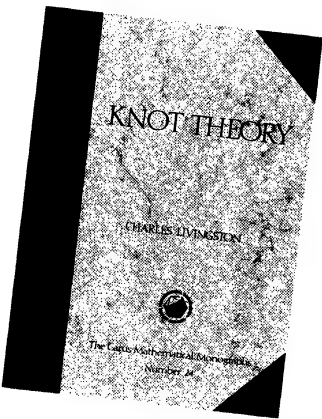
Charles Livingston

I learned more about knots after an hour with the book than I thought I could, and I am glad that it is here on my desk so that I may spend more time with it and, I hope, learn more.
—Paul Halmos

Knot Theory, a lively exposition of the mathematics of knotting, will appeal to a diverse audience from the undergraduate seeking experience outside the traditional range of studies to mathematicians wanting a leisurely introduction to the subject. Graduate students beginning a program of advanced study will find a worthwhile overview, and the reader will need no training beyond linear algebra to understand the mathematics presented.

Over the last century, knot theory has progressed from a study based largely on intuition and conjecture into one of the most active areas of mathematical investigation. **Knot Theory** illustrates the foundations of knotting as well as the remarkable breadth of techniques it employs—combinatorial, algebraic, and geometric.

The interplay between topology and algebra, known as algebraic topology, arises early in the book, when tools from linear algebra and from basic group theory are introduced to study the properties of knots, including the unknotting number, the braid index, and the bridge number. Livingston guides you through a general survey of the topic showing how to use the techniques of linear algebra to address some sophisticated problems, including one of mathematics' most beautiful topics, symmetry. The book closes with a discussion of high-dimensional knot theory and a presentation of some



of the recent advances in the subject—the Conway, Jones and Kauffman polynomials. A supplementary section presents the fundamental group, which is a centerpiece of algebraic topology.

An extensive collection of exercises is included. Some problems focus on details of the subject matter; others introduce new examples and topics illustrating both the wide range of knot theory and the present borders of our understanding of knotting. All are designed to offer the reader the experience and pleasure of working in this fascinating area.

264 pp., Hardbound, 1993

ISBN 0-88385-027-3

List: \$31.50 MAA Member: \$25.00

Catalog Number CAM-24

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036
1-(800) 331-1622 (202)-387-5200



Membership Code

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty. Catalog Number Price

Total \$

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

The Search for E.T. Bell

also known as John Taine

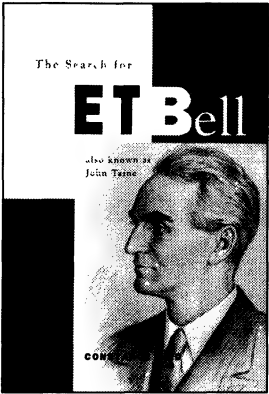
Constance Reid

No one today writes about mathematics and mathematicians with more grace, knowledge, skill, and clarity, and no one is going to produce a more delightful, informative, accurate account of Eric Temple Bell and his work, and that of his alter-ego, the prolific pioneer of science fiction, John Taine. This is a fine book. —Martin Gardner

Eric Temple Bell has been one of my heroes for 60 years...I congratulate Constance Reid on a remarkable achievement. I hope it is greeted with the success it deserves, and revives interest in an extraordinary and multi-talented man. —A. C. Clarke

Eric Temple Bell (1883–1960) was a distinguished mathematician and a best selling popularizer of mathematics. His *Men of Mathematics*, still in print after almost sixty years, inspired scores of young readers to become mathematicians. Under the name “John Taine,” he also published science fiction novels (among them *The Time Stream*, *Before the Dawn*, and *The Crystal Horde*) that served to broaden the subject matter of that genre during its early years.

In *The Search for E.T. Bell*, Constance Reid has given us a compelling account of this complicated, difficult man who never divulged to anyone, not even to his wife and son, the story of his early life and family background. Her book is thus more of a mystery than a traditional biography. It begins with the discovery of an unexpected inscription in an English churchyard and a series of cryptic notations in a boy’s schoolbook. Then comes an inadvertent revelation, by Bell himself, in a respected mathematical journal...You will have to read the book to learn the rest.



Originally agreeing to write only a profile of Bell, Mrs. Reid soon found herself involved in a full-length biography. The discoveries she made in the course of her five years of research will necessitate a fresh evaluation of his extensive mathematical work and his science fiction novels as well as the revision of almost every statement currently in print about his family background and early life. Mrs. Reid is already well known as the author of acclaimed biographies of David Hilbert, Richard Courant, and Jerzy Neyman.

Includes a collection of over 75 photographs.

384 pp., Hardbound, 1993

ISBN 0-88385-508-9

List: \$35.00 MAA Member: \$28.00

Catalog Number BELL

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
(202) 387-5200 1-(800)331-1622

Qty.	Catalog Number	Price
------	----------------	-------

_____	_____	_____
_____	_____	_____

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

Membership Code

Name _____

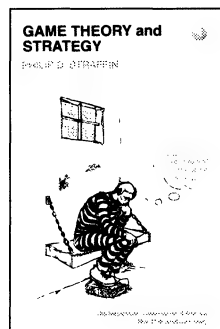
Address _____

City _____

State _____ Zip Code _____

Game Theory and Strategy

Philip D. Straffin, Jr.



This valuable addition to the New Mathematical Library series pays careful attention to applications of game theory in a wide variety of disciplines. The applications are treated in considerable depth. The book assumes only high school algebra, yet gently builds to mathematical thinking of some sophistication. **Game Theory and Strategy** might serve as an introduction to both axiomatic mathematical thinking and the fundamental process of mathematical modelling. It gives insight into both the nature of pure mathematics, and the way in which mathematics can be applied to real problems.

Since its creation by John von Neumann and Oskar Morgenstern in 1944, game theory has contributed new insights to business, politics, economics, social psychology, philosophy, and evolutionary biology. In this book, the fundamental ideas of game theory share the stage with applications of the theory. How might strategic business decisions depend on information about a rival company, and how much would such information be worth? When is it advantageous to vote for a candidate who is not your favorite? What are the optimal strategies for teams in the football draft, and what paradoxes can result from following

those strategies? What is a fair way to share the costs of a development project? What can we learn about the problem of "free will" by imagining playing a game with an omnipotent Being? How might natural selection lead to altruistic behavior in animal species? Game theory gives insight into all of these questions.

The book includes many exercises, with answers, which allow the reader to try out calculations, and explore alternative formulations of game-theoretic ideas.

200 pp., 1993, Paperbound

ISBN 0-88385-637-9

List: \$27.50 MAA Member: \$22.00

Catalog Number NML-36

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
(202) 387-5200 Fax (800) 331-1622



Membership Code

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty. Catalog Number Price

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

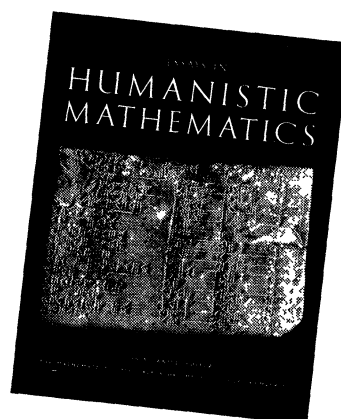
Credit Card No. _____

Signature _____

Exp. Date _____

Essays in Humanistic Mathematics

Alvin White, Editor



A dazzling array of essayists reveal humanistic mathematics in this volume, and in so doing go beyond the facts, formulas, and algorithms that most students associate with mathematics to a presentation of mathematics as an intellectual discipline with a human perspective and a significant history. Humanistic mathematics challenges dogmatic teaching styles that expect students to parrot the lecturer. It demands creativity from both the teacher and student.

Teaching mathematics humanistically seeks to place the student more centrally in the position of inquirer than is generally the case, while at the same time acknowledging the emotional climate of the activity of learning mathematics. This type of teaching encourages students to learn from each other and to better understand mathematics as socially constructed knowledge, rather than as an arbitrary discipline.

Teaching humanistic mathematics brings the focus less upon the nature of the teaching and learning environment and more upon the need to reconstruct the curriculum and the discipline of mathematics itself. This reconstruction relates mathematical discoveries to personal courage, discovery to verification, mathematics to science, truth to utility, and

mathematics to the culture in which it is embedded.

The humanistic mathematics movement, which began as the personal vision of a few, has now become a major part of mathematical culture. What was viewed with skepticism is now accepted and expected. Humanistic mathematics is not a new discovery. It is a recent rediscovery of ideas that go back to Plato. It has provided a vocabulary for previously unarticulated concepts and approaches.

The essays in this volume illustrate and help to define humanistic mathematics. The variety and scope indicate the richness and fruitfulness of the concept. Although each essay is independent, a sense of unity emerges. A glimpse at the table of contents will give you an idea of the excitement and range of the ideas presented.

212 pp., Paperbound, 1993

ISBN 0-88385-089-3

List: \$24.00

Catalog Number NTE-32

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty.	Catalog Number	Price

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

Memorabilia Mathematica

The Philomath's Quotation Book

Robert Edouard Moritz

When Robert Edouard Moritz compiled his book of quotations, **Memorabilia Mathematica**, which appeared in 1914, he stated that his primary objective was to seek out the exact statement of and exact references for famous passages about mathematics. He searched the writing not only of mathematicians, but poets, philosophers, historians, statesmen, and scientists as well. His sources ranged from the works of Plato to the writings of Hilbert and Whitehead. His second objective was to produce a volume that would be a source of pleasure, encouragement, and inspiration to both mathematicians and non-mathematicians alike.

This work was a ten-year labor of love, and it is a tribute to his discerning eye that this selection of passages should remain one of the most stimulating works about mathematics ever published. It was the first collection of its kind in English and it conveys a sense of the full range of mathematics, its enormous accomplishments, and the living personalities of great mathematicians.

The more than eleven-hundred fully annotated selections in this book, gathered from the works of three hundred authors, cover a vast range of subjects pertaining to mathematics. Grouped in twenty-one chapters, they deal with such topics as the definitions and objects of mathematics; the teaching of mathematics; mathematics as a language or as a fine art; the relationship of mathematics to philosophy, to logic, or to science; the

nature of mathematics, and the value of mathematics. Other sections contain passages referring to specific subjects in the field such as arithmetic, algebra, geometry, calculus, and modern mathematics. Of special interest is the extensive amount of material on great mathematicians which provides irreplaceable glimpses into the lives and personalities of mathematical giants.

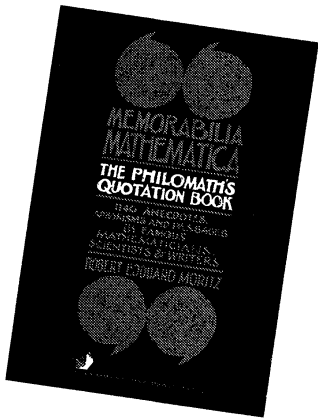
To mathematicians the book will be a great source of pleasure, inspiration, and encouragement. To teachers of mathematics and writers about mathematics, it will remain of inestimable value as a source of quotations and ideas. To the layperson, it will be a revelation. It should dispel forever the narrow notion that mathematics is a cut-and-dried affair, isolated from other compartments of life and thought.

440 pp., Paperbound, 1993

ISBN 0-88385-321-3

List: \$24.00 MAA Member: \$19.00

Catalog Number: MEMO



ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty.	Catalog Number	Price
------	----------------	-------

--	--	--

--	--	--

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

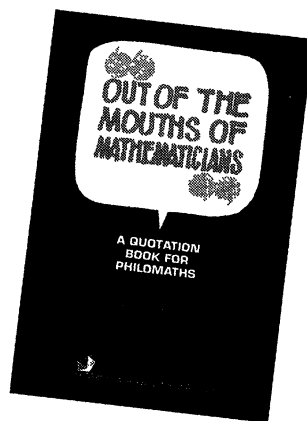
Out of the Mouths of Mathematicians

A Quotation Book for Philomaths

Rosemary Schmalz

Published as a companion volume to Robert Edouard Moritz's **Memorabilia Mathematica**, Rosemary Schmalz's **Out of the Mouths of Mathematicians** picks up where Moritz left off. Her work will give you a sense of the "story" of twentieth century mathematics. You will encounter the mathematicians, their collaborations and disputes, the movement from abstraction to application, the emergence of new areas of research, the impact of computers on mathematics, the challenges in mathematics education, and more.

Out of the Mouths of Mathematicians: A Quotation Book for Philomaths is a compilation of 727 quotations from 292 contributors, almost all of whom are twentieth century mathematicians. Taking the advice of Abel to learn mathematics by reading the masters, the author offers the reader a unique perspective on this century's mathematics through the words of the mathematicians who are its creators. Stories about these mathematicians, their exhortations to their students, their descriptions of their efforts, successes, and failures, all make this century's mathematics come alive. The book also offers readers the opportunity to broaden their ideas about what mathematics is by offering many definitions of mathematics, making comparisons of mathematics to computing and to the fine arts, and showing similarities between many aspects of mathematics and religion. The complete reference for each quotation allows the reader to continue exploration into a favorite area. A large topic index makes the book quite user-friendly. Some of the subject categories include:



The Development of Mathematics, Exhortations to Aspiring Mathematicians, Pure and Applied Mathematics, About Mathematicians (by name), Anecdotes and Miscellaneous Humor, Particular Disciplines in Mathematics, Moments of Mathematical Insight, Mathematics and the Arts,... and much more.

This book will give pleasure to any philomath. It can be used to facilitate a literature search or to give quick access to an appropriate quote for writers and speakers. It will be particularly useful to teachers of mathematics at all levels, to encourage, motivate, and amuse their students. Along with R. E. Moritz's earlier book of this type, **Memorabilia Mathematica: The Philomath's Quotation Book**, it offers the story of mathematics from its primary source, the mathematicians themselves.

304 pp., Paperbound, 1993

ISBN 0-88385-509-7

List: \$29.00 MAA Member: \$23.00

Catalog Number OMMA

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Name _____

Address _____

City _____

State ____ Zip Code _____

Qty.	Catalog Number	Price

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

NEW IN THE SPECTRUM SERIES

Complex Numbers and Geometry

Liang-shin Hahn

The purpose of this book is to demonstrate that complex numbers and geometry can be blended together beautifully, resulting in easy proofs and natural generalizations of many theorems in plane geometry—such as the Napoleon theorem, the Ptolemy–Euler theorem, the Simson theorem, and the Morley theorem.

Beginning with a construction of complex numbers, readers are taken on a 140-page guided tour that includes something for everyone, even those with advanced degrees in mathematics. Yet, the entire book is accessible to a talented high-school student.

The book is self-contained—no background in complex numbers is assumed—and can be covered at a leisurely pace in a one-semester course. Many of the chapters can be read independently. Over 100 exercises are included. The book would be suitable as a text for a geometry course, or for a problem solving seminar, or as enrichment for the student who wants to know more.

200 pp., Paperbound, 1994

ISBN 0-88385-510-0

List: \$25.50 MAA Member: \$19.50

Catalog Number CNGE

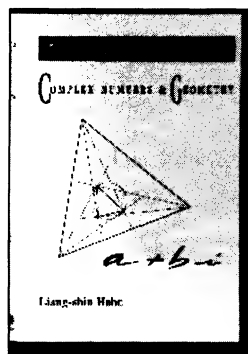


Table of Contents

1. Complex Numbers

Introduction to Imaginary Numbers; Definition of Complex Numbers; Quadratic Equations; Significance of the Complex Numbers; Order Relation in the Complex Field; The Triangle Inequality; The Complex Plane; Polar Representation of Complex Numbers; The n th Roots of 1; The Exponential Functions; Exercises

2. Applications to Geometry

Triangles; The Ptolemy–Euler Theorem; The Clifford Theorems; The Nine-Point Circle; The Simson Line; Generalizations of the Simson Theorem; The Cantor Theorems; The Feuerbach Theorem; The Morley Theorem; Exercises

3. Möbius Transformations

Stereographic Projection; Möbius Transformations; Cross Ratios; The Symmetry Principle; A Pair of Circles; Pencils of Circles; Fixed Points and the Classification of Möbius Transformations; Inversions; The Poincaré Model of a Non-Euclidean Geometry; Exercises

Name _____

Address

City

State Zip Code _____

Qty.	Catalog Number	Price
------	----------------	-------

Total \$

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No.

Signature _____

Exp. Date

NEW IN THE SPECTRUM SERIES

Cryptology

Albrecht Beutelspacher

FR BRX XQFHUVWCQG WKLV? If you can't decipher this coded message, you must read this book!

How can messages be transmitted secretly? How can one guarantee that the message arrives safely in the right hands exactly as it was transmitted? Cryptology—the art and science of “secret writing”—provides ideal methods to solve these problems of data security.

Technology advances have stimulated interest in the study of cryptology. Of course, computers can break cryptosystems much more efficiently than humans can. Computers allow complex and sophisticated mathematical techniques which achieve a degree of security undreamt of by previous generations. Today the applications of cryptology range from the encryption of television programs sent via satellite, to user authentication of computers, to new forms of electronic payment systems using smart cards.

The first half of the book studies and analyzes classical cryptosystems. Here we find Caesar's cipher, the Spartan scytale, the Vigenère cipher, and more. The theory of cipher systems is presented, including a description of the best possible cipher, the one-time pad. An introduction to linear shift registers, which serve as

building blocks for most presently used ciphers, is also given.

The second half of the book looks at the exciting new directions of public-key cryptology, which since its invention in 1976, has revolutionized data security. The author also looks at the famous RSA-algorithm, algorithms based on “discrete logarithms,” the so-called zero-knowledge algorithms, and the smart cards that bring cryptographic services to the man-on-the-street.

Although the mathematics covered is nontrivial, the book is fun to read, and the author presents the material clearly and simply. Many exercises and references accompany each chapter. The book will appeal to a wide audience including teachers, students, and the interested layman.

Cryptology was originally published in German by Vieweg. This edition has been extensively revised.

176 pp., Paperbound, 1994

ISBN 0-88385-504-6

List: \$26.00 MAA Member: \$20.00

Catalog Number CRYPT

Qty. Catalog Number Price

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

Name _____

Address _____

City _____

State _____ Zip Code _____

Proofs Without Words

Exercises in Visual Thinking

Roger B. Nelsen

Just what are “proofs without words?” First of all, most mathematicians would agree that they certainly are not “proofs” in the formal sense. Indeed, the question does not have a simple answer. Proofs without words are generally pictures or diagrams that help the reader see *why* a particular mathematical statement may be true, and *how* one could begin to go about proving it. While in some proofs without words an equation or two may appear to help guide that process, the emphasis is clearly on providing *visual* clues to stimulate mathematical thought. Proofs without words bear witness to the observation that often in the English language to *see* means to *understand*, as in “to see the point of an argument.”

Proofs without words have a long history. In this collection you will find modern renditions of proofs from ancient China, classical Greece, twelfth-century India—even one based on a published proof by a former President of the United States! However, most of the proofs are more recent creations, and many are taken from the pages of MAA journals.

The proofs in this collection are arranged by topic into six chapters: Geometry and Algebra; Trigo-

nometry, Calculus and Analytic Geometry; Inequalities; Integer Sums; Sequences and Series; and Miscellaneous. Teachers will find that many of the proofs in this collection are well suited for classroom discussion and for helping students to think visually in mathematics.

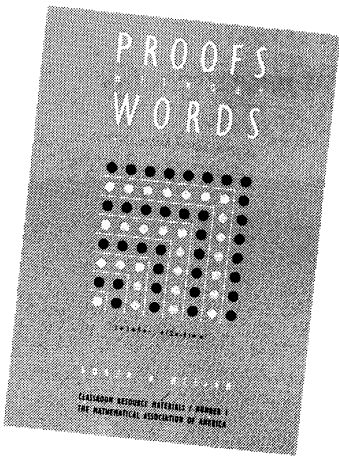
The readers of this collection will find enjoyment in discovering or rediscovering some elegant visual demonstrations of certain mathematical ideas that teachers will want to share with their students. Readers may even be encouraged to create new “proofs without words.”

160 pp., Paperbound, 1993

ISBN 0-88385-700-6

List: \$27.50 MAA Member: \$22.00

Catalog Number PWW



ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty.	Catalog Number	Price
------	----------------	-------

_____	_____	_____
-------	-------	-------

_____	_____	_____
-------	-------	-------

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

EXCURSIONS IN CALCULUS:

an Interplay of the Continuous and the Discrete

Robert M. Young

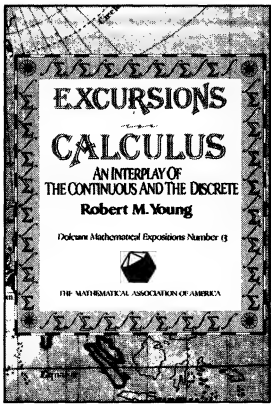
An excellent source of projects for well motivated students. This list of 463 references is a valuable aid for those who wish to dig deeper. —CHOICE

The presentation is clear and the topics very interesting...fully accessible to students for whom the book is intended. The book will be influential in awakening students' awareness for good classical mathematics. —Paulo Ribenboim

Printed with eight full-color plates.

The purpose of this book is to explore, within the context of elementary calculus, the rich and elegant interplay that exists between the two main currents of mathematics, the continuous and the discrete. Such fundamental notions in discrete mathematics as induction, recursion, combinatorics, number theory, discrete probability, and the algorithmic point of view as a unifying principle are continually explored as they interact with traditional calculus. The interaction enriches both.

The book is addressed primarily to well-trained calculus students and their teachers, but it can serve as a supplement in a traditional calculus course for anyone who wants to see more.



CONTENTS:

- Infinite Ascent, Infinite Descent: The Principle of Mathematical Induction
- Patterns, Polynomials, and Primes: Three Applications of the Binomial Theorem
- Fibonacci Numbers: Function and Form
- On the Average
- Approximation: from Pi to the Prime Number Theorem
- Infinite Sums: A Potpourri

The problems, taken for the most part from probability, analysis and number theory, are an integral part of the text. Many point the reader toward further excursions. There are over 400 problems presented in this book.

408 pp., 1992, Paperbound
ISBN 0-88385-317-5
List: \$39.00 MAA Member: \$31.00
Catalog Number DOL-13

ORDER FROM:

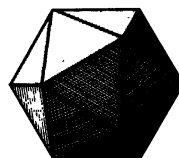
The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

Membership Code	Qty.	Catalog Number	Price

Name _____			
Address _____			
City _____			
State ____ Zip Code _____			
			Total \$ _____
			Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
			Credit Card No. _____
			Signature _____ Exp. Date _____

The American Mathematical Monthly

Volume 101, Number 5 / MAY 1994
(ISSN 0002-9890)



Contents

ARTICLES

On the Geometry of Piecewise Circular Curves / THOMAS BANCHOFF
and PETER GIBLIN 403

The Two Envelope Paradox / ELLIOT LINZER 417

Fourier Series of Polygons / ALAIN ROBERT 420

The Paradox of Nontransitive Dice / RICHARD P. SAVAGE, JR. 429

Squares Expressible as Sum of Consecutive Squares /
LAURENT BEECKMANS 437

Square Roots mod p / STEPHEN M. TURNER 443

FEATURES

COMMENTS 402

NOTES

Kummer's Test Gives Characterizations for Convergence or Divergence
of all Positive Series / JINGCHENG TONG 450

Isometries of ℓ_p -norm / CHI-KWONG LI and WASIN SO 452

A Trace Inequality for Unitary Matrices / BOYING WANG
and FUZHEN ZHANG 453

An Elementary Proof of the Square Summability of the Discrete Hilbert
Transform / LOUKAS GRAFAKOS 456

THE COMPUTER SCIENCE SAMPLER

Does Anybody Really Know What Time It Is? /
CATHERINE C. MCGEOCH 459

THE EVOLUTION OF...

How Hyperbolic Geometry Became Respectable /
ABE SHENITZER 464

THE AUTHORS 471

PROBLEMS AND SOLUTIONS 473

REVIEWS

The Lure of the Integers. By Joe Roberts / PAUL T. BATEMAN
and HAROLD G. DIAMOND 480

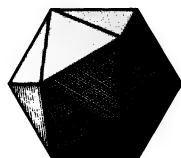
Excursions in Calculus: An Interplay of the Continuous and the Discrete.
By Robert M. Young / ANITA E. SOLOW 482

TELEGRAPHIC REVIEWS 485

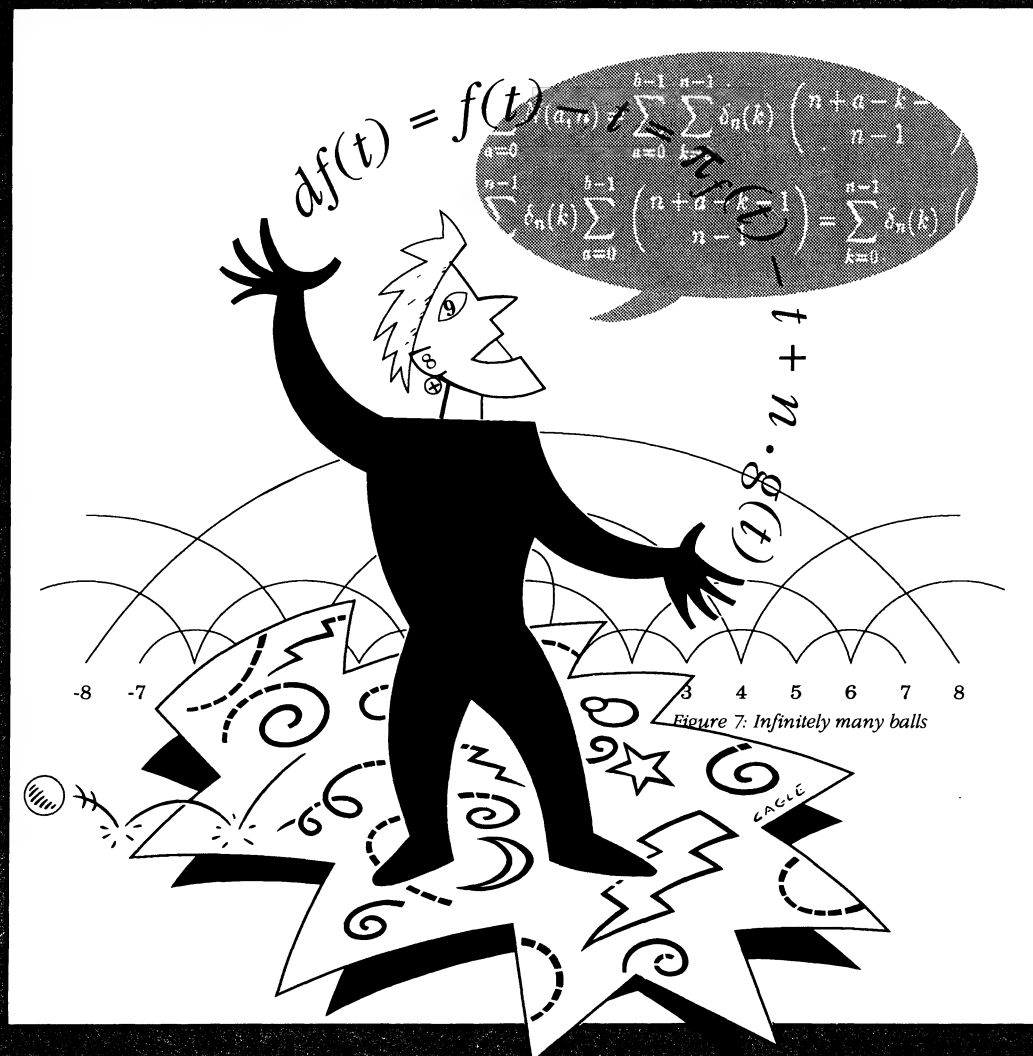
THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.



The American Mathematical Monthly



Volume 101, Number 6 / JUNE-JULY 1994



NOTICE TO AUTHORS

The *Monthly* publishes articles, notes, and other features about mathematics and the profession. The readership of the *Monthly* is intended to include everybody who is mathematically inclined, including of course professional mathematicians and students of mathematics at all collegiate levels. While no single article or feature is likely to appeal to everyone, material should interest and be accessible to a large number of readers. This is the most important criterion for acceptance.

Articles may be expositions of old results or presentations of new ones. They may concern all of mathematics or one small area, a broad development or a single application, historical reminiscences or one important event. While some articles may contain the author's new research, the novelty of material and generality of the results is far less important than the clarity of exposition and general interest. Discussing one illuminating case of a well known result is far better than providing all the details of an obscure but new proposition. Articles in the *Monthly* are supposed to inform and to entertain; they are meant to be read rather than archived.

Notes are short and possibly informal articles. A note may concern a clever new proof of an old theorem, a novel way to present tired material, or a lively discussion of a philosophical (but still mathematical) issue. Also, any topic is suitable, so long as it is related to mathematics. Because a note is short, the first few sentences are the most important part: They should explain the purpose and invite the reader in. Photographs or diagrams often will attract the reader's attention.

All articles and notes should be sent to the editor:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Please send 3 copies, typewritten on only one side of the paper. Illustrations should be carefully drawn on separate sheets of paper in black ink; the original should be without lettering and two copies should have appropriate captions and lettering indicated.

Proposed problems or solutions should be sent to:

RICHARD BUMBY,
P.O. Box 10971
New Brunswick, NJ 08906-0971.

Please send 2 copies of all material, typewritten if possible.

Letters to the Editor, both for publication and for private reading, should be sent to the Editor at the address given above. Comments, including criticisms, are welcome, as are all suggestions for making the *Monthly* a lively, entertaining, and informative journal.

EDITOR:

JOHN H. EWING

ASSOCIATE EDITORS:

PETER BORWEIN	FRED KOCHMAN
RICHARD BUMBY	CATHERINE MCGEOCH
DENNIS DETURCK	RICHARD NOWAKOWSKI
UNDERWOOD DUDLEY	ARNOLD OSTEBEE
JOHN DUNCAN	LEE RUBEL
JOAN FERRINI-MUNDY	ABE SHENITZER
JOSEPH GALLIAN	LYNN STEEN
STEVEN GALOVICH	STAN WAGON
RICHARD GUY	DOUGLAS WEST
DARRELL HAILE	HERBERT WILF
PAUL HALMOS	SANDY ZABELL
JOAN HUTCHINSON	PAUL ZORN

EDITORIAL ASSISTANT:

MISTY CUMMINGS

STAFF ARTIST:

MIKE CAGLE

Reprint permission:

MARCIA P. SWARD, Executive Director

Advertising Correspondence:

Ms. ELAINE PEDREIRA, Advertising Manager

Subscription correspondence, change of address, and other inquiries:

Membership / Subscriptions Department

All at the address:

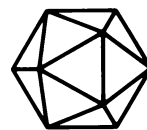
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT. Copyrighted by the Mathematical Association of America (Incorporated), 1994, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source. Second class postage paid at Washington, DC, and additional mailing offices. **Postmaster:** Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.

The American Mathematical Monthly

Volume 101 Number 6 / JUNE–JULY 1994
(ISSN 0002-9890)



Contents

ARTICLES

Juggling Drops and Descents / JOE BUHLER, DAVID EISENBUD,
RON GRAHAM, and COLIN WRIGHT 507

Teaching Integration by Substitution / DAVID GALE 520

Workable Gears, Archimedean Solids and Planar Bipartite Graphs /
GARY GORDON 527

On the Kummer Solutions of the Hypergeometric Equation /
REESE T. PROSSER 535

Reflections on a Mira / JOHN W. EMERT, KAY I. MEEKS,
and ROGER B. NELSON 544

Buffon Noodles / ED WAYMIRE 550

FEATURES

COMMENTS 506

PICTURE PUZZLE 559

NOTES

On a Curious Property of Counting Sequences / VICTOR BRONSTEIN
and AVIEZRI S. FRAENKEL 560

Chaos Without Nonperiodicity / CARSTEN KNUDSEN 563

A Reverse Stolarsky's Inequality / JOSIP PEČARIĆ 565

A Note on Some Irrational Decimal Fractions /
A. MCD. MERCER 567

THE AUTHORS 569

UNSOLVED PROBLEMS

A Possible Permanent Formula / DAVID CALLAN 571

PROBLEMS AND SOLUTIONS 574

REVIEWS

Ideals, Varieties, and Algorithms. By David Cox, John Little,
and Donal O'Shea / MOSS SWEEDLER 582

TELEGRAPHIC REVIEWS 587

Comments

High school mathematics is in a very unhealthy condition at the present time. When a doctor seeks to cure a patient, it is important for him to determine the cause of the malady.

Fifty years ago, mathematics was king of the required subjects; mathematics was supposed to develop, strengthen, and discipline the mind. However, early in the present century, educators and administrators began to question seriously the value of the mathematics then being taught. Teachers with little knowledge of mathematics or rigor were constantly insisting on mathematical rigor. Mortality in mathematics was very high; and as the great influx of pupils of lesser ability began, this became a serious problem. Finally the entire doctrine of formal discipline was swept away . . .

Unfortunately, leaders in mathematics did not first put their house in order before essaying its defense. As a result we have had a forty-year war between the educator and the mathematician, with mathematics constantly on the losing end. Like most wars, this has been unnecessary. More than fifty years ago the great leaders, Klein, Tannery, Perry, Borel, Eliot, Judd, Young, and Moore, pointed the way to reform that would have given us mathematics so easy to defend that no defense would have been necessary.

We can easily double the value of high school mathematics, make it far more interesting to our pupils and much easier to teach by heeding the advice given by E. H. Moore in 1903 and by dozens of other great leaders since his time. . . .

To be brief at the expense of a certain amount of oversimplification, the reforms needed are as follows:

1. We must stop trying to teach mathematics by the stupid "water-tight compartment" method . . . W. D. Reeve says, "Our traditional 'water-tight compartment' method of teaching algebra, then geometry, then intermediate algebra, leads to a great deal of unnecessary repetition of subject matter that results in the loss of time and energy."

2. We must remove the large amount of relatively useless material that is found in our best modern texts, and introduce more interesting and more practical mathematics to take the place of the dead wood eliminated. More than twenty-five years ago a committee of the Association reported: ". . . The present standards of drill work, largely on non-essentials, were set up about fifty years ago. A considerable number of teachers . . . believe that the amount of time spent by pupils on abstract work in difficult problems in division, factoring, fractions, simultaneous equations, radicals, et cetera, is excessive; that such work leads to nothing important in the science."

3. We must have ability grouping in mathematics. There must be at least two and at best three types of levels of mathematics in high schools. . . . We have always had ability grouping in athletics. If the results in the teaching of mathematics were as immediately determined as in athletics, we should have ability grouping at once. Ability grouping is the essence of true democracy, that is, each pupil entitled to be developed to his limit without being held back by less capable pupils. . . .

If the problem can be solved, it will be worth many millions of dollars yearly to the country and will be one of the most important projects in the history of American education. To secure the needed and long overdue reforms in mathematics we must have a committee of experts with adequate financing, so that they may devote their full time to this most important project. The final report should be printed and given wide distribution. Part of this report could be a set of specimen or model textbooks . . .

— C. N. Shuster, MONTHLY 55(1948), 472-475

Juggling Drops and Descents

Joe Buhler, David Eisenbud, Ron Graham, and Colin Wright

As circus and vaudeville performers have known for a long time, juggling is fun. In the last twenty years or so this has led to a surge in the number of amateur jugglers. It has been observed that scientists, and especially mathematicians and computer scientists, are disproportionately represented in the juggling community. It is difficult to explain this connection in any straightforward way, but music has long been known to be popular among scientists; juggling, like music, combines abstract patterns and mind-body coordination in a pleasing way. In any event, the association between mathematics and juggling may not be as recent as it appears, since it is believed that the tenth century mathematician Abu Sahl started out juggling glass bottles in the Bagdad marketplace ([3], p. 79).

In the last fifteen years there has been a corresponding increase in the application of mathematical and scientific ideas to juggling ([1], [2], [7], [11], [13], [18]), including, for instance, the construction of a juggling robot ([8]). In this article we discuss some of the mathematics that arises out of a recent juggling idea, sometimes called “site swaps.” It is curious that these idealized juggling patterns lead to interesting mathematical questions, but are also of considerable interest to “practical” jugglers. The basic idea seems to have been discovered independently by a number of people; we know of three groups or individuals that developed the idea around 1985: Bengt Magnusson and Bruce Tiemann ([12], [11]), Paul Klimek in Santa Cruz, and one of us (C. W.) in conjunction with other members of the Cambridge University Juggling Association. A precursor of the idea can be found in [14].

Although our interests here are almost entirely mathematical, the reader interested in actual juggling or its history might start by looking at [21] and [19]; a leisurely discussion of site swaps, aimed at jugglers, can be found in [12].

In the first section we describe the basic ideas, and in the second section we prove the basic combinatorial result that counts the number of site swaps with a given period and a given number of balls. This theorem has a non-obvious generalization to arbitrary posets ([6]). Special cases of that result can be interpreted in terms of an interesting generalization of site swaps; we find it delightful that a question arising from juggling leads to new mathematics which in turn may say something about patterns that jugglers might want to consider.

1. JUGGLING. As mathematicians are in the habit of doing, we start by throwing away irrelevant detail. In a juggling pattern we will ignore how many people or hands are involved, ignore which objects are being used, and ignore the specific paths of the thrown objects. We will assume that there are a fixed number of objects (occasionally referred to as “balls” for convenience) and will pay attention

only to the times at which they are thrown, and will assume that the throw times are periodic. Although much of the interest of actual juggling comes from peculiar throws (behind the back, off the head, etc.), peculiar objects (clubs, calculus texts, chain saws, etc), and peculiar rhythms, we will find that the above idealization is sufficiently interesting.

Suppose that you are juggling b balls in a constant rhythm. Since the throws occur at discrete equally-spaced moments of time, and since in our idealized world you have been juggling forever and will continue to do so, we identify the times t of throws with integers $t \in \mathbf{Z} := \{\dots, -2, -1, 0, 1, 2, \dots\}$.

Since it would be silly to hold onto a ball forever, we assume that each ball is thrown repeatedly. We also assume that only one ball is thrown at any given time. With these conventions, a juggling pattern with b balls is described, for our purposes, by b doubly-infinite disjoint sequences of integers.

The three ball cascade is perhaps the most basic juggling trick. Balls are thrown alternately from each hand and travel in a figure eight pattern. The balls are thrown at times

ball 1:	$\dots -6, -3, 0, 3, 6, \dots$
ball 2:	$\dots -5, -2, 1, 4, 7, \dots$
ball 3:	$\dots -4, -1, 2, 5, 8, \dots$

This pattern has a natural generalization for any odd number of balls; if you tried to do this pattern with an even number of balls (in a symmetrical way) then two balls would collide at the middle of the figure eight.

Another basic pattern, sometimes called the fountain or waterfall, is most commonly done with an even number of balls and consists of two disjoint circles of balls.

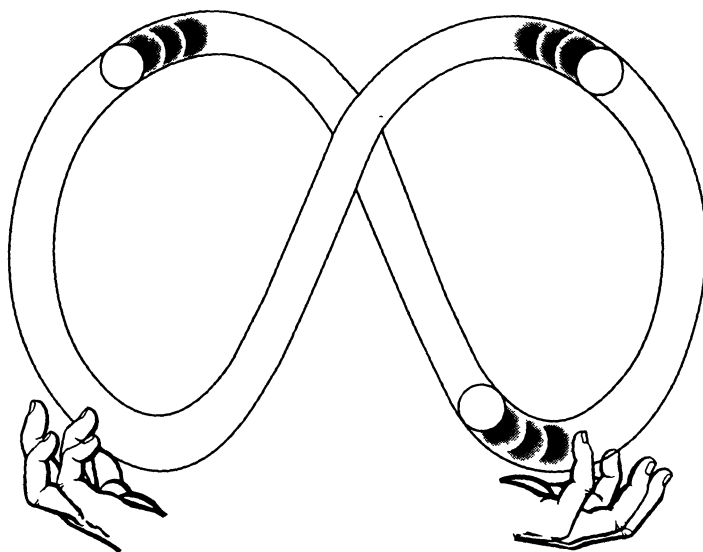


Figure 1. A cascade.

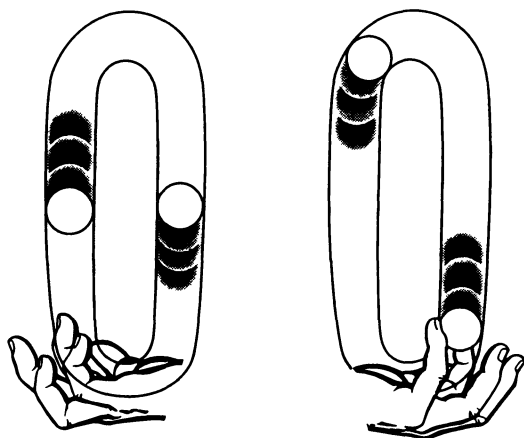


Figure 2. A fountain.

The four ball waterfall gives rise to the four sequences $\{4n + a: n \in \mathbb{Z}\}$ of throw times, for $a = 0, 1, 2, 3$.

The last truly basic juggling pattern is called the shower. In a shower the balls travel in a circular pattern, with one hand throwing a high throw and the other throwing a low horizontal throw. The shower can be done with any number of balls; most people find that the three ball shower is significantly harder than the three ball cascade. The three ball shower corresponds to the sequences

$$\begin{aligned} \text{ball 1:} & \quad \dots -6, -5, 0, 1, 6, 7 \dots \\ \text{ball 2:} & \quad \dots -4, -3, 2, 3, 8, 9 \dots \\ \text{ball 3:} & \quad \dots -2, -1, 4, 5, 10, 11 \dots \end{aligned}$$

We should mention that although non-jugglers are often sure that they have seen virtuoso performers juggle 17 or 20 balls, the historical record for a sustained ball cascade seems to be nine. Enrico Rastelli, sometimes considered the greatest juggler of all time, was able to make twenty catches in a 10-ball waterfall pattern. Rings are somewhat easier to juggle in large numbers, and various people have been able to juggle 11 and 12 rings.

Now we return to our idealized form of juggling. Given lists of throw times of b balls define a function $f: \mathbb{Z} \rightarrow \mathbb{Z}$ by

$$f(x) = \begin{cases} y & \text{if the ball thrown at time } x \text{ is next thrown at time } y \\ x & \text{if there is no throw at time } x. \end{cases}$$

This function is a permutation of the integers. Moreover, it satisfies $f(t) \geq t$ for all $t \in \mathbb{Z}$. This permutation partitions the integers into orbits which (ignoring the orbits of size one) are just the lists of throw times.

The function $f(t) = t + 3$ corresponds to the 3-ball cascade, which could be graphically represented as in FIGURE 4.

Similarly, the function $f(t) = t + 4$ represents the ordinary 4-ball waterfall. The three ball shower corresponds to a function that has a slightly more complicated description. The juggler is usually most interested in the duration $f(t) - t$

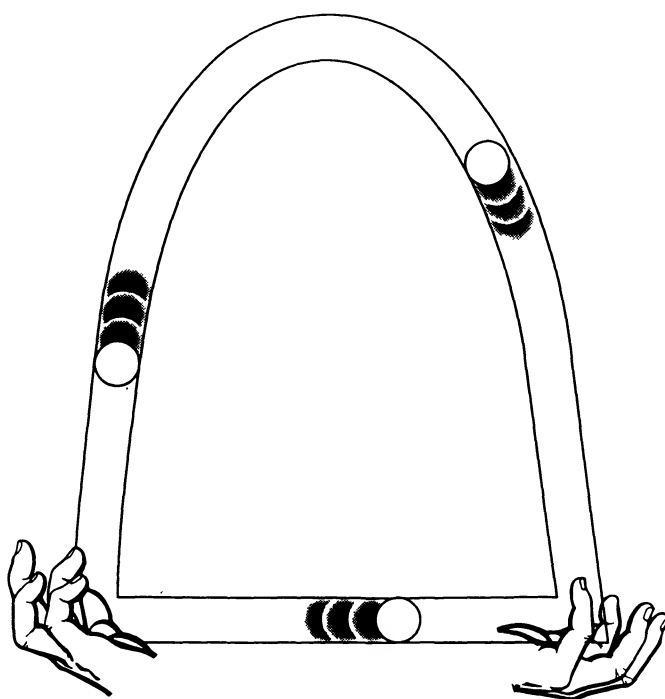


Figure 3. A shower.

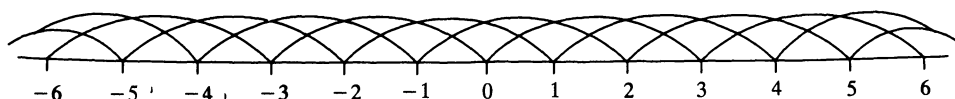


Figure 4. $t \rightarrow t + 3$.

between throws which corresponds, roughly, to the height to which balls must be thrown.

Definition. A *juggling pattern* is a permutation $f: \mathbf{Z} \rightarrow \mathbf{Z}$ such that $f(t) \geq t$ for all $t \in \mathbf{Z}$. The *height function* of a juggling pattern is $df(t) := f(t) - t$.

The three ball cascade has a height function $df(t) = 3$ that is constant. The three ball shower has a periodic height function whose values are $\dots 5, 1, 5, 1, \dots$. The juggling pattern in FIGURE 5 corresponds to the function

$$f(x) = \begin{cases} x + 4 & \text{if } x \equiv 0, 1 \pmod{3} \\ x + 1 & \text{if } x \equiv 2 \pmod{3} \end{cases}$$

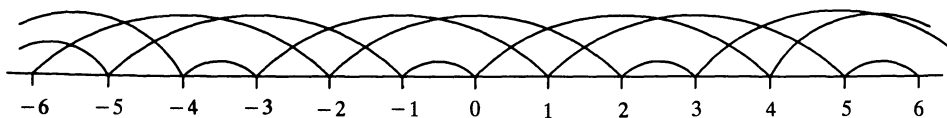


Figure 5. 441

which is easily verified to be a permutation. The height function takes on the values 4, 4, 1 cyclically. This trick is therefore called the “441” among those who use the standard site swap notation. It is not terribly difficult to learn but is not a familiar pattern to most jugglers.

Remarks:

1. We refer to $df(t)$ as the height function even though it more properly is a rough measure of the elapsed time of the throw. From basic physics the height is proportional to the square of the elapsed time. The elapsed time is actually less than $df(t)$ since the ball must be held before being thrown. For a more physical discussion of actual elapsed times and throw heights see [11].
2. Although there is nothing in our idealized setup that requires two hands, or even “hands” at all, we note that in the usual two-handed juggling patterns, a throw with odd throw height $df(t)$ goes from one hand to the other, and a throw with even throw height goes from one hand to itself.
3. If $f(t) = t$, so that $df(t) = 0$, then no throw takes place at time t . In actual practice this corresponds to an empty hand.
4. Nothing in our model really requires that the rhythm of the juggling pattern be constant. We only need a periodic pattern of throw times. We retain the constant rhythm terminology in order to be consistent with jugglers’ standard model of site swaps.
5. The catch times are irrelevant in our model. Thus a throw at time t of height $df(t)$ is next thrown at time $t + df(t) = f(t)$, but in practice it is caught well before that time in order to allow time to prepare for the next throw. A common time to catch such a throw is approximately at time $f(t) - 1.5$ but great variation is possible. A theorem due to Claude Shannon ([13], [7]) gives a relationship between flight times, hold times, and empty times in a symmetrical pattern.

Now let f be a juggling pattern. This permutation of \mathbf{Z} partitions the integers into orbits; since $f(t) \geq t$, the orbits are either infinite or else singletons.

Definition. The number of balls of a juggling pattern f , denoted $B(f)$, is the number of infinite orbits determined by the permutation f .

Our first result says that if the throw height is bounded, which is surely true for even the most energetic of jugglers, then the number of balls is finite and can be calculated as the average value of the throw heights over large intervals.

Theorem 1. *If f is a bijection and $df(t) = f(t) - t$ is non-negative and bounded then the limit*

$$\lim_{|I| \rightarrow \infty} \frac{\sum_{x \in I} df(x)}{|I|}$$

exists and is equal to $B(f)$, where the limit is over all integer intervals

$$I = \{a, a + 1, \dots, b\} \subset \mathbf{Z}.$$

Proof: Suppose that $df(t) \leq B$ for all t . If I is an interval such that $|I| > B$ then any infinite orbit intersects I . The sum of $df(t)$ over the points in I lying in a given infinite orbit is bounded above by $|I|$ and below by $|I| - 2B$. If I is large enough

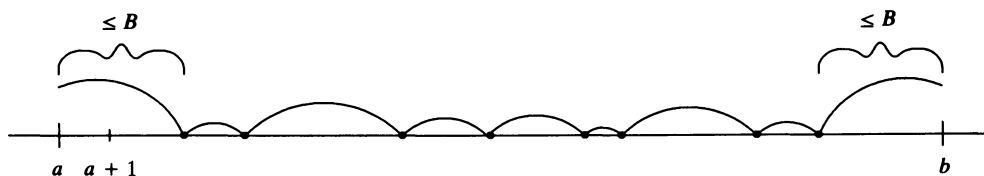


Figure 6. One orbit.

then the sum of $df(t)$ for $t \in I$ can be made arbitrarily close to the number of infinite orbits of f ; the singleton orbits don't contribute since $df(t) = 0$ for those orbits. Thus in the limit the average of df over an interval $\{a, a+1, \dots, b\}$ of consecutive integers must become arbitrarily close to the number of infinite orbits of the permutation. ■

Remarks:

1. The limit is clearly a uniform limit in the sense that for all positive ε there is an m such that if I is an interval of integers with more than m elements then the average of df over I is within ε of $B(f)$.
2. As an example illustrating the theorem we note if f is the 441 pattern described earlier, then the height function $df(t)$ is periodic of period 3. The long term average of $df(t)$ over any interval approaches the average over the period, i.e., $(4 + 4 + 1)/3 = 3$, which confirms what we already knew: the 441 pattern is a 3-ball trick.
3. The hypothesis of bounded throw heights is necessary. Indeed, if $T(0) = 0$ and, for nonzero t , $T(t)$ is the highest power of 2 that divides t then the pattern $f(t) = t + 2 \cdot T(t)$ has unbounded throw height and infinite $B(f)$, as in FIGURE 7. More vividly: you can juggle infinitely many balls if you can throw arbitrarily high.

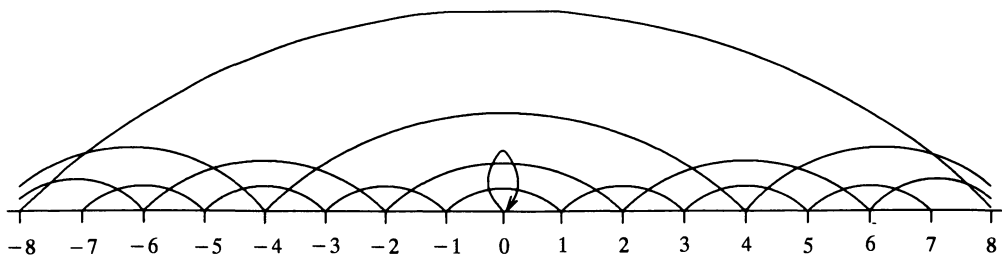


Figure 7. Infinitely many balls.

2. PERIODIC JUGGLING. From now on we want to juggle periodically. A juggling pattern is perceived to be periodic by an audience when its height function is periodic in the mathematical sense.

Definition. A *period- n juggling pattern* is a bijection $f: \mathbf{Z} \rightarrow \mathbf{Z}$ such that $df(t+n) = df(t)$ for all $t \in \mathbf{Z}$.

If df is of period n then it might also have a period m for some divisor m of n . If n is the smallest period of df then any other period is a multiple of n ; in this case we will say that f is a pattern of **exact** period n .

A period- n juggling pattern can be described by giving the finite sequence of non-negative integers $df(t)$ for $t = 0, 1, \dots, n - 1$. Thus the pattern 51414 denotes a period-5 pattern; by Theorem 1 it is a 3-ball pattern since the “period average” of the height function $df(t)$ is 3.

Which finite sequences correspond to juggling patterns? Certainly a necessary condition is that the average must be an integer. However this isn’t sufficient. The sequence 354 has average 3 but does not correspond to a juggling pattern—if you try to draw an arrow diagram for a map f as above you’ll find that no such map exists. This is also easy to see directly, for if $df(1) = 5$ and $df(2) = 4$ then

$$f(1) = 1 + df(1) = 6 = 2 + df(2) = f(2)$$

and such a map isn’t a bijection.

Remarks for Jugglers Only

1. The above description is geared towards the standard model: two hands throwing alternately, in constant rhythm. In fact there could be any number of hands and it is not necessary to assume that the rhythm is constant.
2. The practical meaning of the throw heights 0, 1, and 2 in the standard model requires a little thought. A throw height of 0 corresponds to an empty hand. A throw height of 1 corresponds to a rapid shower pass from one hand to another that is thrown again immediately. A throw height of 2 would ordinarily indicate a very low throw from a hand to itself that is thrown again by that hand immediately. This is actually rather unnatural in practice; the conventional interpretation ([11], [12]) is that a throw height of 2 is a held ball.
3. The paradigm for categorizing juggling patterns here is very interesting in practice, although many of the patterns require considerable proficiency. Several jugglers who have spent time in working on site swaps describe the same gain in flexibility and conceptual power that mathematicians seem to report from the use of well-chosen abstractions. The simplest non-obvious site-swap seems to be 441; it is similar to, but **not** the same as, the common 3-ball pattern of throwing balls up on the side while passing a ball back and forth underneath in a shower pass from hand to hand. The 3-ball 45141 pattern is also amusing, and the 4-ball 5551 pattern looks very much like the 5-ball cascade. The range of feasible and interesting tricks seems to be unlimited; we mention the following sample: 234, 504, 345, 5551, 40141, 561, 633, 55514, 7562, 7531, 566151, 561, 663, 771, 744, 753, 426, 459, 9559, 831.
4. A number of programs are available that simulate site swaps on a computer screen, sometimes with quite impressive graphics. These programs take a finite sequence of non-negative integers as input and dynamically represent the pattern. The Internet news group rec.juggling is a source of information on site swaps and various juggling animation software.

In order to find out which finite sequences represent juggling patterns we start by noting that a period- n pattern induces a permutation on the first n integers.

Lemma. *If f is a period- n juggling pattern then*

$$s \equiv t \pmod{n} \Rightarrow f(s) \equiv f(t) \pmod{n}.$$

Proof: If $df(t)$ is periodic of period n then the function $f(t) = t + df(t)$ is of period n modulo n . ■

The Lemma implies that a juggling pattern f induces a well-defined injective, and hence bijective, mapping on the integers modulo n . Let $[n]$ denote the set $\{0, 1, \dots, n-1\}$ and let S_n denote the symmetric group of all permutations (bijections) of the set $[n]$. Then for every period- n juggling pattern f there is a well-defined permutation $\pi_f \in S_n$ that is defined by the condition

$$f(t) \equiv \pi_f(t) \pmod{n}, \quad 0 \leq t < n.$$

Theorem 2. *A sequence $a_0 a_1 \cdots a_{n-1}$ of non-negative integers satisfies $df(t) = a_t$ for some period- n juggling pattern f if and only if $a_t + t \pmod{n}$ is a permutation of $[n]$.*

Proof: Suppose that f is a juggling pattern and $a_t = df(t)$. Then $f(t) \equiv \pi_f(t) \pmod{n}$ so there is an integer-valued function $g(t)$ such $f(t) = \pi_f(t) + n \cdot g(t)$ and

$$df(t) = f(t) - t = \pi_f(t) - t + n \cdot g(t)$$

and

$$a_t + t \equiv df(t) + t \equiv \pi_f(t) \pmod{n}$$

and $a_t + t$ is a permutation as claimed.

Conversely, suppose that

$$a_0 a_1 \cdots a_{n-1}$$

is such that $a_t + t$ is a permutation of $[n]$. If we define a_t for all integers t by extending the sequence periodically and then define $f(t) = a_t + t$ then f is the desired juggling pattern. To see that f is injective note that if $f(t) = f(u)$ then $t \equiv u \pmod{n}$ since $f(t)$ is injective modulo n . Then $a_t = a_u$. From $f(t) = a_t + t = f(u) = a_u + u$ it follows that $t = u$ and f is injective as claimed. To show that f is surjective, suppose that $u \in \mathbb{Z}$. Since $t + a_t \pmod{n}$ is a permutation of $[n]$ we can find a t such that $f(t) = t + a_t \equiv u \pmod{n}$. By adding a suitable multiple of n we can find a t' such that $f(t') = u$. This finishes the proof of the fact that any sequence satisfying the stated condition comes from a juggling pattern. ■

To see if 345 corresponds to a juggling pattern we add t to the t -th term and reduce modulo 3. The result is 021, which is a permutation, so 345 is indeed a juggling pattern (in fact a somewhat difficult one that is quite amusing). On the other hand, the sequence 354 leads, by the same process, to 000 which certainly isn't a permutation of $[3]$.

Let $N(b, n)$ denote the number of period- n juggling patterns f with $B(f) = b$. Our goal is to calculate this number. From the juggler's point of view it might be more useful to count the number of patterns of exact period n and to count cyclic shifts of a pattern as being essentially the same as the original pattern. Later we will see that this more natural question can be answered easily once we know $N(b, n)$.

The basic idea in the determination of $N(b, n)$ is to fix a permutation $\pi \in S_n$ and count the number of patterns f such that $\pi_f = \pi$. From the proof of the previous theorem we have the formula

$$f(t) = \pi_f(t) + n \cdot g(t) = \pi(t) + n \cdot g(t), \quad 0 \leq t < n.$$

Thus we must count the number of functions $g: [n] \rightarrow \mathbf{Z}$ such that if f is defined by the above formula then $df(t) \geq 0$ and $B(f) = b$.

The number of balls of such a pattern f is equal to the average of $df(t)$ over $[n]$. Thus

$$B(f) = \frac{1}{n} \sum_{t=0}^{n-1} df(t) = \frac{1}{n} \sum_{t=0}^{n-1} (\pi(t) - t + n \cdot g(t)).$$

Since $\pi(t)$ is a permutation of $[n]$ we see that this reduces to

$$B(f) = \sum_{t=0}^{n-1} g(t).$$

Thus a function g determines a pattern with $B(f) = b$ if the sum of its values is equal to b .

The condition that $df(t) \geq 0$ is a little bit more intricate. Since

$$df(t) = \pi(t) - t + n \cdot g(t)$$

we see that $g(t)$ must be non-negative and also must be strictly positive whenever $\pi(t) < t$.

Definition. An integer $t \in [n]$ is a drop for the permutation $\pi \in S_n$ if $\pi(t) < t$; moreover, we define

$$d_\pi(t) = \begin{cases} 1 & \text{if } t \text{ is a drop for } \pi \\ 0 & \text{if } t \text{ is not a drop for } \pi. \end{cases}$$

Write $G(t) = g(t) - d_\pi(t)$ so that

$$f(t) = \pi(t) + n \cdot d_\pi(t) + n \cdot G(t).$$

Let k be the number of drops of π . Then $B(f) = b$ if and only if the sum of the values of G is equal to $b - k$.

We can summarize this discussion so far as follows. The number $N(b, n)$ of period- n juggling patterns with b balls is equal to the sum over all permutations $\pi \in S_n$ of the number of non-negative functions $G(t)$ on $[n]$ whose value-sum is $b - k$, where k is the number of drops of π .

A standard combination idea can be used to count the number of sequences of non-negative integers with a given sum.

Lemma. *The number of non-negative n -tuples with sum x is*

$$\binom{x + n - 1}{n - 1}.$$

Proof: A standard “stars and bars” argument (in Feller’s terminology, e.g., p. 38 of [9]) gives the answer. The number of such sequences is equal to the number of ways of arranging $n - 1$ bars and x stars in a row if we interpret the size of each contiguous sequence of stars as a component of the n -tuple and the bars as separating components. The number of such sequences of bars and stars is the same as the number of ways to chose $n - 1$ locations for the bars out of a total of $x + n - 1$ locations, which is just the stated binomial coefficient. ■

Let $\delta_n(k)$ be the number of permutations in S_n that have k drops. By combining the earlier remark with the lemma we arrive at

$$N(b, n) = \sum_{k=0}^{n-1} \delta_n(k) \binom{n+b-k-1}{n-1}.$$

Later it will be convenient to consider the number of period- n juggling patterns with fewer than b balls. If this number is denoted $N_{<}(b, n)$ then, using a familiar binomial coefficient identity, we find that

$$\begin{aligned} N_{<}(b, n) &= \sum_{a=0}^{b-1} N(a, n) = \sum_{a=0}^{b-1} \sum_{k=0}^{n-1} \delta_n(k) \binom{n+a-k-1}{n-1} \\ &= \sum_{k=0}^{n-1} \delta_n(k) \sum_{a=0}^{b-1} \binom{n+a-k-1}{n-1} = \sum_{k=0}^{n-1} \delta_n(k) \binom{n+b-k-1}{n-1}. \end{aligned}$$

In order to simplify this further we recall the idea of a descent of a permutation and show that even though drops and descents aren't the same thing, the number of permutations with k drops is the same as the number with k descents.

Definition. If $\pi \in S_n$ then $i \in [n]$ is a **descent** of π if $\pi(i) < \pi(i+1)$ where $0 \leq i < n-1$. The number of elements of S_n with k descents is denoted

$$\left\langle \begin{matrix} n \\ k \end{matrix} \right\rangle$$

and is called an **Eulerian number**.

We will write permutations as a list of n integers in which the i -th element is $\pi(i)$, e.g.,

$$\pi(0)\pi(1) \dots \pi(n-1).$$

A descent in π is just a point in this finite sequence in which the next term is lower than the current term.

Example. The permutation 10432 in S_5 has three descents and two drops.

If π is a permutation then it can also be written in cycle form in the usual way. In order to specify this form uniquely we write each cycle with its largest element first and arrange the cycles so that the leading elements of the cycles are in increasing order, where we include the singleton cycles.

Definition. If $\pi \in S_n$ let $\hat{\pi}$ be the permutation that results from writing π in cycle form, as above, and then erasing parentheses.

Example. The permutation $\pi \in S_8$ corresponding to the sequence 16037425 has a cycle decomposition (0162)(475) that has the canonical form (3)(6201)(765). Therefore $\hat{\pi}$ is 36201754.

Note that the map taking π to $\hat{\pi}$ is bijective since π can be uniquely reconstructed from $\hat{\pi}$ by inserting left parentheses before every left-to-right maximum and then inserting matching right parentheses. This permutation of S_n is certainly bizarre at first glance, but it plays a surprisingly crucial role in various situations (see [5] or [15]).

Lemma. *The number of permutations of $[n]$ with k descents is equal to the number with k drops, i.e.,*

$$\delta_n(k) = \left\langle n \atop k \right\rangle.$$

Proof: A descent of $\hat{\pi}$ must lie inside a cycle of π since our conventions guarantee that the last element in a cycle is followed by a larger integer. By the meaning of the cycle decomposition π (namely, that elements within cycles are mapped to the next element in the cycle) we see that a descent of $\hat{\pi}$ corresponds to a drop of π . Conversely, a drop in π must occur within a cycle (i.e., not in passing from the last element of a cycle to the first) and corresponds to a descent in $\hat{\pi}$. Thus the number of permutations with k descents is equal to the number $\delta_n(k)$ with k drops. ■

Example, again. The permutation $\pi = 16037425$ has drops at $t = 2, 5, 6, 7$, and the permutation $\hat{\pi} = 36201754$ has descents at $i = 1, 2, 5, 6$.

The Eulerian numbers $\delta_n(k) = \left\langle n \atop k \right\rangle$ play a role in a variety of combinatorial questions beyond drops and descents ([10], [15], [16]), although no notation seems to be standard yet. We recall some of their basic properties. If a permutation $\pi = \pi(0)\pi(1)\dots\pi(n-1)$ has k descents then its reversal $\pi' = \pi(n-1)\pi(n-2)\dots\pi(0)$ has $n-k-1$ descents. Thus

$$\left\langle n \atop k \right\rangle = \left\langle n \atop n-k-1 \right\rangle. \tag{1}$$

By relating permutations of $[n]$ to permutations of $[n-1]$ in the usual way, a more involved combinatorial argument shows that

$$\left\langle n \atop k \right\rangle = (k+1)\left\langle n-1 \atop k \right\rangle + (n-k)\left\langle n-1 \atop k-1 \right\rangle. \tag{2}$$

Using this recursion, it is easy to tabulate Eulerian numbers; for $n \leq 5$ they are given in Table 1.

TABLE 1. Eulerian Numbers

k					
n	0	1	2	3	4
1	1				
2	1	1			
3	1	4	1		
4	1	11	11	1	
5	1	26	66	26	1

Finally, the Eulerian numbers arise as coefficients of the linear relations connecting the polynomials x^n with the polynomials $\binom{x+k}{n}$.

Worpitzky's Identity.

$$x^n = \sum_{k=0}^{n-1} \left\langle n \atop k \right\rangle \binom{x+k}{n}.$$

This identity can be readily proved by induction using equation (2). It apparently first appeared in [20] (see also [10] and [16]); in [15] it appears as a special case of a much more general statement.

Theorem 3. *The number of period- n juggling patterns with fewer than b balls is b^n , i.e.,*

$$N_{<}(b, n) = b^n.$$

Proof: Our previous formula for $N_{<}(b, n)$ was

$$N_{<}(b, n) = \sum_{k=0}^{n-1} \delta_n(k) \binom{n+b-k-1}{n} = \sum_{k=0}^{n-1} \left\langle \begin{matrix} n \\ k \end{matrix} \right\rangle \binom{n+b-k-1}{n}.$$

Replace k by $n-k-1$ and use (2) to get

$$N_{<}(b, n) = \sum_{k=0}^{n-1} \left\langle \begin{matrix} n \\ k \end{matrix} \right\rangle \binom{b+k}{n}.$$

The claim is then an immediate consequence of Worpitzky's identity. ■

The simplicity of the final result is surprising. The astute reader will note that we could have avoided introducing the concept of descents by proving equations (1) and (2) directly for the counting function $\delta_n(k)$ for drops. It is a pleasant exercise to provide a direct combinatorial argument. We took the slightly longer route above because it is amusing and useful in proving the more general result in [6].

By the theorem there are $(b+1)^n - b^n$ patterns of period n with exactly b balls if cyclic shifts are counted as distinct. Let $M(n, b)$ be the number of patterns of exact period n with exactly b balls, where cyclic shifts are not counted as distinct. Thus $M(n, b)$ is probably the number that is of most interest to a juggler.

If d is a divisor of n then each pattern of exact period d will occur d times as pattern of length n . Thus

$$(b+1)^n - b^n = \sum_{d|n} dM(d, b).$$

By Möbius inversion we obtain the following corollary to the previous theorem.

Corollary.

$$M(n, b) = \frac{1}{n} \sum_{d|n} \mu(n/d) ((b+1)^d - b^d).$$

For instance, there are 12 genuinely distinct patterns with period three with three balls. The reader may find it instructive to list all of them explicitly.

Several people have reproved Theorem 3 from other points of view. Richard Stanley sent us a proof using results in [15]. Jeremy Kahn sent us a bijective proof using a different labeling function for juggling patterns. Walter Stromquist sent us an interesting bijective proof that uses a very curious relabeling of site swap patterns. Adam Chalcraft ([4]) sent us a proof using ideas similar to those of Stromquist. It is striking that the result seems to be of considerable interest to a number of people.

Several of these proofs are shorter than ours, and some are much closer to being more transparent “bijective” proofs. However, the proof given here, in addition to using some interesting combinatorics, is the special case of the proof of the more general result in [6]. The basic motivation of that result is to replace the set $[n]$ with an arbitrary poset. For some posets we can give a natural interpretation of that more general result in terms of juggling patterns in which more than

one ball can be thrown at once, but we still haven't been able to give a juggling interpretation for arbitrary posets. After hearing of our results from Richard Stanley, E. Steingrímsson reproved ([17]) the general results about posets using results from his thesis. Among many other things, he generalizes the notions of descents and drops (actually, in his terminology, a mirror notion he calls "exceedances") to certain wreath products of symmetric groups.

NOTE ADDED IN PROOF: In their recent preprint, "Juggling and applications to q -analogues," Richard Ehrenborg and Margaret Readdy give a q -analogue of our main result. In addition they generalize the ideas to multiplex patterns (in which a hand can catch and throw more than one ball at once) and give applications to q -Stirling numbers and the Poincaré series of an affine Weyl group.

REFERENCES

1. H. Austin, *A Computational View of the Skill of Juggling*, M.I.T. Artificial Intelligence Laboratory, 1974.
2. P. J. Beek, *Juggling Dynamics*, Free University Press, Amsterdam, 1989.
3. J. L. Berggren, *Episodes in the Mathematics of Medieval Islam*, Springer Verlag, 1986.
4. A. Chalcraft, manuscript in preparation.
5. D. Bayer and P. Diaconis, Trailing the Dovetail Shuffle to its Lair, Technical Report, Department of Statistics, Stanford, 1989.
6. J. Buhler and R. Graham, A note on the drop polynomial of a poset, in preparation.
7. J. Buhler and R. Graham, Fountains, showers, and cascades, *The Sciences*, Jan.-Feb. 1984, 44-51.
8. M. Donner, A real-time juggling robot, IBM research preprint.
9. W. Feller, *Introduction to Probability Theory and its Applications*, 3rd edition, John Wiley & Sons, 1968.
10. R. Graham, D. Knuth, and O. Patashnik, *Concrete Mathematics*, Addison Wesley Co., 1989.
11. B. Magnusson and B. Tiemann, The physics of juggling, *Physics Teacher*, 27 (1989) 584-589.
12. B. Magnusson and B. Tiemann, A notation for juggling tricks, *Juggler's World*, summer 1991, 31-33.
13. C. Shannon, Scientific Aspects of Juggling, unpublished manuscript.
14. C. Simpson, Juggling on paper, *Juggler's World*, winter 1986, 31.
15. R. Stanley, *Enumerative Combinatorics*, Wadsworth & Brooks/Cole, 1986.
16. D. Stanton, *Constructive Combinatorics*, Springer-Verlag, 1986.
17. E. Steingrímsson, Permutation statistics of indexed and poset permutations, Ph.D. dissertation, MIT, 1991.
18. B. Summers, Juggling as performing mathematics, *Co-Evolution Quarterly*, summer 1980.
19. M. Truzzi, On keeping things up in the air, *Natural History*, 1979, 44-55.
20. J. Worpitzky, Studien über die Bernoullischen und Eulerschen Zahlen, *Journal für die reine und angewandte Mathematik*, 94 (1881) 103-232.
21. K.-H. Ziethen and A. Allen, *Juggling, The Art and its Artists*, Werner Rausch & Werner Lüft Inc., 1985.

Buhler:
 Reed College
 Portland, OR 97202
 jpb@reed.edu

Graham:
 AT&T Bell Laboratories
 Murray Hill, NJ 07974
 rlg@research.att.com

Eisenbud:
 Brandeis University
 Waltham, MA 02254
 eisenbud@math.brandeis.edu

Wright:
 The University of Liverpool
 Liverpool, L69 3BX, England

Teaching Integration by Substitution

David Gale

The current boom in calculus reform programs has been going on now for more than six years at a cumulative cost of well over five million dollars. A major theme of the program has been the need to get away from so-called cook book calculus, to teach concepts rather than techniques, understanding rather than rote memorization. This is of course a worthy goal but just how one goes about achieving it is not at all obvious. What I want to do in the paragraphs which follow is to look at this question in the context of a special case by treating a particular calculus question which has been bothering me on and off for more than 50 years.

To begin at the beginning, I took freshman calculus in 1940 and in the intervening years I have taught virtually that same course to others dozens of times. I found the course back then mildly disappointing in that it seemed to consist for the most part of working hundreds of drill problems, but I managed to master enough of the tricks to struggle through with a grade of B. There was one thing in the course, though, that really bothered me, and that was the matter of integration by substitution. I didn't have any trouble with integrands like $x\sqrt{1-x^2}$ and $\ln(x)/x$. I understood that in general one hopes that the integrand will have the form $f(g(x))g'(x)$, in which case if one happens to know an antiderivative of f , call it F , then the antiderivative one is looking for is $F(g(x))$. It was also clear, as the book explained, that in fact this was nothing but the chain rule in reverse. But then one day we had to integrate $\sqrt{1-x^2}$ without the extra x on the outside, so the book, "Calculus" by Arnold Dresden, said, well, make the substitution $x = \sin(t)$. Then $dx = \cos(t) dt$ etc. etc. Again, I could go through the mechanics without difficulty but this time it seemed to me the operations were not justified by anything we had done up to that time. In the earlier cases we made a substitution of the form $u = g(x)$ but now we were supposed instead to write $x = g(t)$, which didn't seem to me to be the same thing.

Because of the current interest in calculus instruction I decided now, after more than half a century it would be interesting to see how textbooks these days are handling the substitution problem that had thrown me off as a student. To this end I looked in 10 fairly traditional texts, some of which are among the current best sellers. The authors are Anton (to be abbreviated A), Edwards and Penny (EP), Ellis and Gulick (EG), Lang (L), Larson and Hostetler (LH), Marsden and Weinstein (MW), Stein (SN), Stewart (ST), Swokowski (SW), Thomas and Finney (TF). Also I looked at a draft of an as yet unpublished text by the Harvard Calculus Consortium (H). Here are some of my findings.

All of these books use and prove the "direct" substitution theorem devoting an entire section to the subject. None of them proves what I will call the inverse substitution theorem although all of them except L and H use it fairly intensively, devoting a whole section to Trigonometric Substitution. Only ST explicitly recognizes the fact that inverse substitution is not the same as direct substitution, and I

will return to this in a moment. I then looked at how the direct substitution theorem was treated, and all of the books did indeed show how it followed from the chain rule. In six of the books, however, A, FP, L, MW, ST and TF, I found the following formula. If $u = g(x)$ then

$$\int f(g(x))g'(x) dx = \int f(u) du. \quad (1)$$

Of course the equation is false. The expression $\int f(x) dx$ stands for antiderivative, as in a table of integrals, and the variable, be it x , t , u or anything else is a dummy. Clearly the antiderivatives on the left and right above are not equal. What the books mean, no doubt, is that if you substitute $g(x)$ for u after taking the antiderivative on the right you get the antiderivative on the left. I expect some readers will say I am being pedantic or that there is no need to be so rigorous at the freshman level, but I think this kind of lapse is symptomatic of a rather strange set of standards and perhaps it sheds light on why none of the books proves the inverse substitution theorem. It is because none of them formulates it. Once one does, the proof becomes more or less mechanical and one sees at once that it is not a mere application of the chain rule but involves other things, as we will see in a moment. The book ST is interesting. It uses equation (1) above to describe substitution while *inverse substitution* [the book's italics] is described by $x = g(t)$ and

$$\int f(x) dx = \int f(g(t))g'(t) dt \quad (2)$$

where g is required to have an inverse. Notice that (1) and (2) are the same (false) equation. Only the letters for the (dummy) variables are different (of course these equations become correct when one considers definite integrals and puts in the appropriate limits).

Let us now turn to the mathematics which to my surprise turned out to be rather interesting. There are (at least) two different proofs of the inverse substitution theorem. The first is direct (brute force), and slightly messy. The second, suggested to me by Ole Hald, is short and elegant and makes use of some "theory". For the sake of cleanness I will use the circle notation for composition of functions. People who prefer the more traditional $f(g(x))$ notation should have no trouble translating the argument below, at the cost of having to carry around masses of parentheses and a lot of, in my view, superfluous x 's. (The subject of notation, which is rather interesting, will be considered in an appendix.) Differentiation will be denoted by a prime, $'$, and for typographical clarity I will use a dot, \cdot , for multiplication of functions. Finally I will use a block, \blacksquare , for "proof" as well as for "Q. E. D.", a notational reform I have been trying to persuade the mathematical community to adopt for the past 25 years with no success whatsoever. First, then, we have the chain rule,

$$(f \circ g)' = (f' \circ g) \cdot g'. \quad (\text{Ch})$$

Now, both the substitution rules described in the preceding paragraphs deal with the situation where we have three functions h , f and g and

$$h = (f \circ g) \cdot g'. \quad (*)$$

In the *direct substitution* case we know an antiderivative for f and want to find one for h . The answer is given by,

Direct Substitution Theorem. If $F' = f$ then

$$(F \circ g)' = h. \quad (3)$$

■ From the chain rule, $(F \circ g)' = (F' \circ g) \cdot g' = (f \circ g) \cdot g' = h$. ■

In the *inverse substitution* case we know an antiderivative of h and want to find one for f .

Inverse Substitution Theorem. *If $H' = h$ and g has an inverse then*

$$(H \circ g^{-1})' = f. \quad (4)$$

Let me suggest at this point that the reader take two minutes to work out the direct proof of (4) in order to see what is involved.

As one might expect, one needs to use not only the chain rule but also the formula for the derivative of the inverse of a function which in our notation is

$$(g^{-1})' = 1/(g' \circ g^{-1}). \quad (\text{Inv})$$

■ From (Ch) and (*) we have

$$(H \circ g^{-1})' = (H' \circ g^{-1}) \cdot (g^{-1})' = (((f \circ g) \cdot g') \circ g^{-1}) \cdot (g^{-1})'.$$

We must now simplify the term on the right hand side and we need several facts. The first is the general but not so familiar identity that for any functions a , b and c

$$(a \cdot b) \circ c = (a \circ c) \cdot (b \circ c). \quad (5)$$

The right hand side then becomes

$$((f \circ g) \circ g^{-1}) \cdot (g' \circ g^{-1}) \cdot (g^{-1})',$$

which by the associative law for composition and the fact that $g \circ g^{-1}$ is the identity function, simplifies to

$$f \cdot (g' \circ g^{-1}) \cdot (g^{-1})',$$

but the product on the right is 1 by application of (Inv). ■

There is an interesting observation to be made about this proof. It makes no direct assumptions about the function f , not even continuity, but ends up *proving* that f is integrable (meaning antidifferentiable). The second proof, which will now be given, requires the antidifferentiability of f as a hypothesis. Thus, it assumes that h has an antiderivative H , and f has an antiderivative F , but it does not need to make use of (Inv).

■ Since $F' = f$ we have by (Ch), $(F \circ g)' = (f \circ g) \cdot g' = h$, so H and $F \circ g$ have the same derivative and hence (theory) they differ by a constant, thus,

$$F \circ g = H + c. \quad (6)$$

Now composing both sides of (6) on the right with g^{-1} gives

$$F \circ g \circ g^{-1} = F = H \circ g^{-1} + c \circ g^{-1} = H \circ g^{-1} + c,$$

so $(H \circ g^{-1})' = F' = f$. ■

In addition to the standard textbooks above, I looked at some nonstandard ones, those of De Leeuw, Spivak and Moise (there is also the book by Menger but it requires essentially learning a new language so I did not consider it). Believe it or not, the book of Moise actually states and proves the inverse substitution theorem very much as in the first proof above. I could find no other book that does

this. De Leeuw has an entire short section entitled *Integration by Inverse Substitution* but he leaves the theorem itself as an exercise, which I think is a rather good idea and I will elaborate on this at the end. Spivak's book is supposed to be a completely rigorous calculus. He presents one example in which he lets $x = \sin u$ and says "this really means that we are using the substitution $u = \arcsin x$ "—but he does not give the details. (Spivak and De Leeuw both point out, by the way, that equation (1) is not to be taken literally but is merely intended to be "suggestive").

I will return to the inverse substitution theorem in a moment, but first I want to record some further findings of my quickie textbook survey. The overwhelming impression one gets is that of uniformity, both "vertical" and "horizontal". Over the 50 plus years I could detect no general trend toward any kind of change among the standard texts. As for the current crop, if one looks at their tables of contents they seem almost identical up to permutation of chapters and sections. Thus "Trigonometric Substitution" is either Section 5 or 6 of the chapter "Techniques of Integration" and is invariably preceded by a section on Integrals Involving Trigonometric Functions. Integration by Substitution is around Section 5 of the chapter which introduces the integral. Indeed, the whole calculus catechism seems to have become quite rigidly codified. I wonder if this is a strictly national phenomenon. Are French, Russian, Chinese calculus texts very much like ours or are we perhaps teaching "American Calculus"? Maybe one of the 35 NSF grantees has looked into this question.

I want now finally to relate the foregoing discussion to the question of teaching concepts instead of techniques. First, I should say that I believe the inverse substitution theorem is important and should be part of any calculus course, not just because it enables one to find antiderivatives of square roots of quadratic polynomials but more importantly because it is an example of the change of variable technique which is fundamental in much of mathematics. My proposal is to try to induce understanding this material by giving the students a fairly intensive workout on composition of functions, a subject which they often seem to find difficult. The following sequence of problems followed by annotations in brackets is intended to illustrate how this might be done.

Exercises

1. Let $f(x) = 2x + 1$, $g(x) = x - 2$.

- (a) Calculate $f \circ g$ and $g \circ f$.

[Routine but illustrates noncommutativity of composition.]

- (b) Find functions h and k such that $f \circ h = g$ and $k \circ f = g$.

[Not so routine but students will learn something by figuring these out.]

2. (a) Show that the linear function $f(x) = ax + b$ has a linear *inverse*, provided $a \neq 0$, that is, there is a linear function which we denote by f^{-1} such that $(f^{-1} \circ f)(x) = x$. Write out the function f^{-1} .

[This is intended to reinforce the idea that arithmetic calculations are often made with letters rather than numbers.]

- (b) From (a) calculate $(f \circ f^{-1})(x)$.

[In contrast to 1 (b), the inverse of a function *does* commute with the function.]

- (c) If $g(x) = cx + d$ use (a) and (b) to find h and k such that $f \circ h = g$ and $k \circ f = g$. (Hint. Try $h = f^{-1} \circ g$.)

[Again to illustrate the difference between right and left composition.]

3. Let f be any function. Use the chain rule to find the equation obtained by differentiating both sides of the equations

$$(f \circ f^{-1})(x) = x, \quad (f^{-1} \circ f)(x) = x.$$

[This will be needed in the last problem.]

4. Let $f(x) = 2x + 1$ and $g(x) = x^2 - 1$.

(a) Calculate $f \circ g$ and $g \circ f$.

[Routine but a little more complicated.]

(b) Find h and k such that $f \circ h = g$ and $k \circ f = g$.

[Corresponding to 1(b) at the next level.]

(c) Is there any polynomial h such that $g \circ h = f$? Explain.

[One should demand a clearly expressed argument here.]

5. Let $f(x) = 2x^2 + 1$, $g(x) = x^2 - 2$.

(a) Find k such that $k \circ f = g$.

[Somewhat more interesting computation.]

(b) Show that there is no polynomial h such that $f \circ h = g$.

[Again the students should be required to write out the argument in a precise, grammatical manner. If there are TA sections it would be a good idea to ask students to write down their answers on the board.]

5. Let $f(x) = 2x + 1$, $g(x) = x - 2$, $h(x) = 5$.

Calculate $f \circ (g \cdot h)$, $(f \circ g) \cdot h$, $f \cdot (g \circ h)$, $(f \cdot g) \circ h$.

[The answers are $10x - 19$, $10x - 15$, $6x + 3$, 33 . This is to illustrate the importance of where you put the parentheses.]

6. Let f , g and h be *any* functions. Answer True or False and give reasons,

(a) $(f \cdot g) \circ h = (f \circ h) \cdot (g \circ h)$,

[It is perhaps too much to ask for a precise written argument here but the instructor could give the details in class.]

(b) $(f \circ g) \cdot h = (f \cdot h) \circ (g \cdot h)$.

[A counter example is needed. Some of the better students will perhaps find one. Again a good problem for classroom discussion.]

7. Prove the Inverse Substitution Theorem: If $h = (f \circ g) \cdot g'$ and g has an inverse and H is an antiderivative of h then $H \circ g^{-1}$ is an antiderivative of f . You will need to use Problem 3 and Problem 6 (a).

Let me end by summarizing. I have tried (A), to make a case for including a proof of the inverse substitution theorem in calculus courses and, (B), to suggest how this might be done in a way that will lead students to a better understanding of what the result means. My approach is based on the proposition that what students learn in a course is not what they read in a text or hear in a lecture but what they do on their own. Routine questions and drill problems definitely have their place, but if we want the students to understand why these routines work then we must also design problems whose solution will require this understanding. The exposition presented here was an attempt in this direction.

Appendix. Some Thoughts on Notation. The reader may have noticed that in Exercises 2 and 3 above I shamelessly abandoned my principles and wrote $(f \circ f^{-1})(x)$ despite the earlier somewhat derogatory remarks about using unnecessary parentheses and x 's. Also, in the first proof of the Inverse Substitution Theorem I said that something followed because " $g \circ g^{-1}$ is the identity function". How many calculus texts have you seen that talk about the identity function? These difficulties go right to the root of our pedagogical problem, I believe. Namely, I trust everyone agrees that the most important concept in calculus, and probably in all of mathematics, is that of function. Further, it is surely a statistical fact that the most used letter in mathematical discourse is x , yet this letter is consistently used for two entirely different objects, first, variables as in, for example "Let x be the distance...." and second the identity function, as in "the derivative of x is 1". Then there's c , the "constant" which is both a number, as in the "constant of integration", and a function as in "the derivative of a constant is 0". This last fact which in some of the books is listed as the first rule of differentiation is often written

$$dc/dx = 0$$

which is especially troubling. What has x got to do with c ?

I realize, of course, that these observations have been made many times, and I expect some will say that, yes, there is some ambiguity here but we all know from the context what is intended, and fussing about such matters is pedantry. I agree, *we* know what is intended from a lifetime of experience but, remember, our students are seeing these things for the first time. If our mission is, as we claim, to induce understanding then it's hard to see how a deliberately ambiguous notation will help. Furthermore, the remedy is simple. We should give functions their own type face, just as we already do for vectors. Thus, we have the number c and the constant function \mathbf{c} or the number 5 and the function $\mathbf{5}$. The first rule of differentiation is then either

$$\mathbf{c}' = 0 \quad \text{or} \quad D(\mathbf{c}) = 0$$

depending on one's choice of notation,—but please, no x 's.

I suppose one could denote the identity function by \mathbf{x} but I personally have a strong preference for \mathbf{i} . Then in Exercise 3 above the expression $\mathbf{f} \circ \mathbf{f}^{-1} = \mathbf{i}$ looks just fine, and the "second rule of differentiation" is

$$D(\mathbf{i}) = 1$$

and then it's on to

$$D(\mathbf{i}^n) = n\mathbf{i}^{n-1},$$

and so on.

But now there are some debatable points. Consistency requires that we write polynomials as, e.g. $3\mathbf{i}^2 - 2\mathbf{i} + 5$, which does seem a bit radical. I do feel, though, that choice of notation may have a definite influence on how we teach and hence what the students learn.

Perhaps the real reason we can't switch to a sensible notation is that all the current books do things in the traditional way, that we were brought up that way ourselves by our predecessors and therefore this is how we are condemned to pass things on to our successors. But are we really locked in forever to this way of operating? Now that mathematicians have gotten these things sorted out explicitly, (a comparatively recent achievement), one would hope that we could at some point pass along this enlightenment to our students.

There are, I admit, some complications of a practical nature in all of this. In the age of word processing there is no problem in getting all the type faces we need to distinguish numbers, functions, vectors or anything else. The trouble is, though, that as long as lectures are still given the old fashioned way, with chalk at a blackboard, one does not have so many choices. This may explain traditional usage where we assign different parts of the Latin alphabet to different concepts. Thus, a, b, c are constants whereas x, y, z are variables. There are probably psychological reasons for this. The more rarely used, “distant” letters suggest uncertainty or “variability”. Next, f, g, h are functions, which makes sense since f is for function, but why are i and j almost invariably the letters of choice for subscripts? Is it perhaps because i is for index? Of course, m and n are also used for subscripts, but note that they are also the run away favorites for exponents, whereas i and j are rarely used for this purpose, and if someone were to write about Euclidean j -space we probably wouldn’t know what was intended.

Returning finally to my mini textbook survey, I noted a curious phenomenon. Almost all of the books at some point would use two different notations for the same object in a single formula. Here for example is the way the chain rule appears in H.

$$d/dx(f(g(x))) = f'(g(x)) \cdot g'(x)$$

with d/dx on the left and primes on the right. In A one has a slightly different mixture.

$$d/dx[f(u)] = f'(u)du/dx,$$

and in EP

$$D[f(g(x))] = f'(g(x)) \cdot g'(x).$$

In general the practice seems to be to use d/dx or D or D_x on the left and primes on the right. An even more peculiar practice was to use two different notations for composition. In EG, L, MW and TF one reads

$$(f \circ g)'(x) = f'(g(x))g'(x)$$

thus, circle on the left, parentheses on the right. No book that I have seen writes the chain rule consistently in circle notation as I have done above.

Continuing to backtrack, I found that the practice of using primes on the right and other notations on the left occurs not only for the chain rule but for the other differentiation formulas as well, in A, EP, LH and H. For instance, in H the sum rule is given by

$$d/dx(f(x) + g(x)) = f'(x) + g'(x)$$

but the product rule is simply

$$(fg)' = fg' + f'g.$$

Why does this happen? One certainly would not do such things in writing a research paper. Is it because of some sort of pedagogical theory or is it just carelessness? Maybe this could be a “subject for future research”.

*Department of Mathematics
University of California
Berkeley, CA 94720
gale@math.berkeley.edu*

Workable Gears, Archimedian Solids and Planar Bipartite Graphs

Gary Gordon

1. INTRODUCTION. My young daughters have a popular children's game which involves setting up a system of interlocking gears. Each of the plastic gears can be fastened to a plastic board in a variety of places. The gears come in three sizes and can be adorned with colorful plastic animals and other decorations. This toy can hold their attention for a long time (where 'long time' is defined as longer than five minutes). We will call an arrangement of the gears (which consists of a finite number of gears, some pairs of which interlock and some which don't) *workable* if whenever any one gear is turned, all the gears turn freely. Children usually discover that it is quite easy to construct an unworkable arrangement of the gears.

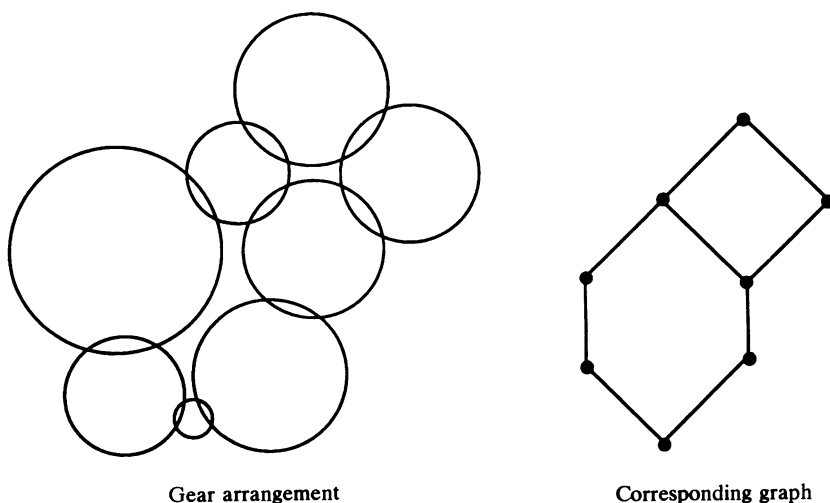


Figure 1

These gears also have an easy and direct connection to graph theory. (We will assume familiarity with most of the basics of graph theory. See [1], [3], [5] or [7] for very readable introductions which emphasize different aspects of the subject.) Given an arrangement S of the gears, define a graph $G(S)$ whose vertices correspond to the gears themselves. Define the edges of $G(S)$ so that uv is an edge joining vertices u and v if and only if the gears corresponding to u and v actually interlock. (See FIGURE 1, where the gears are drawn as circles with interlocked gears overlapping.) Recall that the chromatic number of a graph is the smallest number of colors needed to color the vertices of G so that adjacent

vertices receive different colors. I occasionally give the following question as a homework problem in graph theory courses.

Problem A. *Let $G (= G(S))$ be a graph obtained from a workable arrangement of gears. What can you say about G ? In particular, can you determine the chromatic number of G ?*

Students realize quickly that one thing they can say is that G must be planar, that is, G can be drawn in the plane so that edges don't cross. It's not much harder to see that G will correspond to a workable arrangement of gears if and only if G has no odd cycles. (If the arrangement is workable, then it is clear that G has no odd cycles; the other direction can be established by induction on the number of gears, for example.) Since the students have already learned (or are about to learn) that a graph G has no odd cycles if and only if G is bipartite (i.e., G has chromatic number 2), they can prove the following proposition (and solve Problem A).

Proposition 1. *If G corresponds to a workable arrangement of gears, then G is a planar bipartite graph.*

The rest of this paper is devoted to constructing various planar bipartite graphs which are 'nice' in some way. All of the graphs we will consider are assumed to be connected. In Section 2, we discuss *regular* planar bipartite graphs, i.e., planar bipartite graphs in which every vertex has the same degree. Section 3 is concerned with *semi-regular* planar bipartite graphs, i.e., graphs in which all the red vertices have degree a and all the blue vertices have degree b for given positive integers a and b . These graphs will be closely related to the Archimedean solids. (In a bipartite graph, we will refer to the vertex partition induced by the 2-coloring by simply saying 'the red vertices' or 'the blue vertices'.)

2. REGULAR PLANAR BIPARTITE GRAPHS. If G is a (connected) planar bipartite graph which is regular, then it is easy to show that the number of red vertices and the number of blue vertices must be equal.

Problem B. *Determine all possible positive integers r and n such that there is a planar bipartite graph G which is regular of degree r and which has n red vertices (and n blue vertices).*

The main tool used in solving all of the problems considered here is Euler's famous polyhedral formula, which he discovered around 1750. (See Theorem 8.1.1 of [3], for example.)

Theorem 2 (Euler's Polyhedral Formula). *If a plane drawing of a connected graph with v vertices and e edges has r regions, then $v - e + r = 2$. (This formula includes the unbounded region in the count for r .)*

The next corollary (which appears as Theorem 8.1.5 in [3]) is also a standard result. It follows from Euler's formula and the fact that each cycle in a bipartite graph contains at least four edges.

Corollary 3. *If G is a planar bipartite graph with $v \geq 3$ vertices and e edges, then $e \leq 2v - 4$.*

To solve Problem B, we begin by applying Corollary 3 to G . Let G be a planar bipartite graph which is regular of degree r on $2n$ vertices (n of which are in each of the two color classes). Then the number of edges is given by $e = rn$, so we immediately get $rn \leq 2(2n) - 4$, so $r < 4$.

It remains to investigate the cases $r = 1$, $r = 2$ and $r = 3$ separately. The determination of all possible values for n for the cases $r = 1$ and $r = 2$ is left to the reader (see Proposition 4). We will now determine all possible values of n for the case $r = 3$.

A graph which is regular of degree 3 is called *cubic*. The smallest cubic bipartite graph is $K_{3,3}$, the ‘three houses and three utilities’ graph. (The *complete bipartite graph*, denoted $K_{m,n}$, is the bipartite graph having m red vertices and n blue vertices with every red vertex adjacent to every blue vertex.) $K_{3,3}$ is not planar, so the smallest cubic planar bipartite graph must have $n \geq 4$.

Finding examples of cubic planar bipartite graphs is not hard. We now construct two classes of cubic planar bipartite graphs, one for even n and one for odd n . If $n \geq 4$ is even, define a graph B_n with vertices $\{1, 2, \dots, 2n\}$ as follows: Form two n -cycles, one with vertices $\{1, \dots, n\}$ and the other with vertices $\{n+1, \dots, 2n\}$, then add n edges by joining vertices k and $n+k$ for each k , $1 \leq k \leq n$. (B_4 is isomorphic to the graph associated with a three-dimensional cube.) For odd $n > 5$, define B_n by modifying B_{n-1} as follows: Delete edges $(1, n+1)$, $(3, n+3)$ and $(5, n+5)$ from B_{n-1} , then add two new vertices x and y , joining vertex x to vertices 1, 3 and 5 and vertex y to vertices $n+1$, $n+3$ and $n+5$. See FIGURE 2 for planar drawings of B_6 and B_7 . The red vertices are labeled 1, the blue ones are labeled 2.

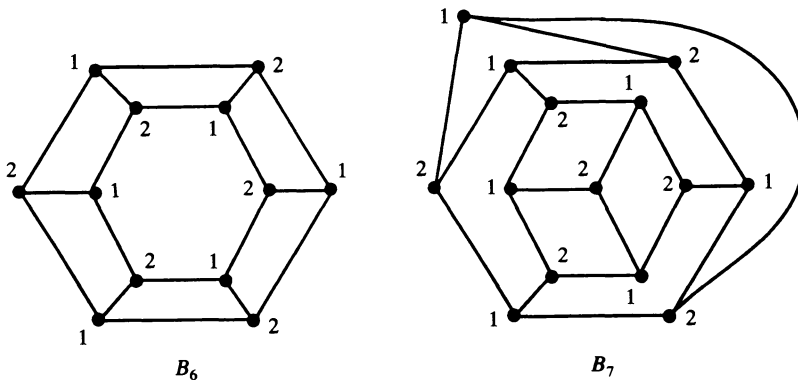


Figure 2

The above procedure fails when $n = 5$. In fact, there is no cubic planar bipartite graph with $n = 5$. If such a graph G existed, then it would have 10 vertices, 15 edges and 7 regions. For a region R , let $e(R)$ denote the number of edges which bound R . Then $e(R)$ is even (since G is bipartite) and $\sum e(R) = 30$ (since each edge is counted twice), where this sum extends over all 7 regions of G . Thus, exactly one of the regions is bounded by a 6-cycle and the six remaining regions are all bounded by 4-cycles. The reader can now show that G cannot be planar; we omit the remaining details.

The next proposition summarizes the results of this section and completely solves Problem B.

Proposition 4. *If G is a (connected) planar bipartite graph which is regular of degree r and has n vertices in each color class, then the only possibilities for n and r are the following:*

- (a) $r = 1$ and $n = 1$,
- (b) $r = 2$ and $n \geq 2$,
- (c) $r = 3$ and $n \geq 4$, with $n \neq 5$.

Furthermore, each of these possibilities can be realized.

3. SEMI-REGULAR PLANAR BIPARTITE GRAPHS AND ARCHIMEDIAN SOLIDS. When we allow a little more flexibility in our planar bipartite graphs, we can get some more interesting examples. In this section we consider semi-regular planar bipartite graphs.

Problem C. *Determine all possible positive integers a and b (with $a < b$) such that there is a semi-regular planar bipartite graph G in which every red vertex has degree a and every blue vertex has degree b .*

Suppose G is a semi-regular planar bipartite graph with m red vertices, each of degree a and n blue vertices, each of degree b , where $a < b$. Then G has $v = m + n$ vertices and $e = am = bn$ edges. By Corollary 3, $am \leq 2m + 2n - 4$. Substituting $m = bn/a$ and simplifying gives the inequality

$$\frac{1}{a} + \frac{1}{b} \geq \frac{1}{2} + \frac{2}{e}. \quad (\ddagger)$$

Thus, the only possible values for a and b are the following:

- 1. $a = 1$ and $b > 1$,
- 2. $a = 2$ and $b > 2$,
- 3. $a = 3$ and $b = 4$ or 5 .

We again leave to the reader the construction of the appropriate graphs for the first two cases above and turn our attention to the case where $a = 3$ and $b = 4$. In the smallest possible example, the inequality in (\ddagger) will be replaced by equality. Solving for the number of edges then gives $e = 24$, so $m = 8$ and $n = 6$ (since $3m = 4n = e$). From Euler's formula, G must have 12 regions. Furthermore, each region of (any planar drawing of) G is bounded by exactly 4 edges. We can now construct G by modifying the graph B_6 constructed above (see Figure 2). As in the modification which produced B_7 , we again add two new vertices and six new edges to B_6 , but this time we don't delete any old edges (see FIGURE 3). We denote this graph by $C_{6,8}$.

We can create an infinite family of semi-regular planar bipartite graphs with $a = 3$ and $b = 4$ by modifying the graphs B_{6k} (for any $k \geq 1$) in an analogous way. Add $2k$ vertices to B_{6k} and join each of these new vertices to three 'blue' vertices of B_{6k} . We denote this family by $C_{6k,8k}$ for $k \geq 1$. See FIGURE 3 for a drawing of $C_{12,16}$, which is a modification of B_{12} .

Since we are considering semi-regular graphs, it is not surprising to discover a connection between the graphs we have constructed and another family of semi-regular graphs, the Archimedean solids. An *Archimedean*, or *semi-regular* solid has the property that its faces are all regular polygons, but of two or more kinds, all its vertices are identical and it can be circumscribed by a regular tetrahedron so that four of its faces lie on the four faces of the tetrahedron. (Dropping the last requirement concerning the circumscribed tetrahedron allows infinite families of prisms and antiprisms and the *pseudo rhombicuboctahedron*. See Figures 2.10 and

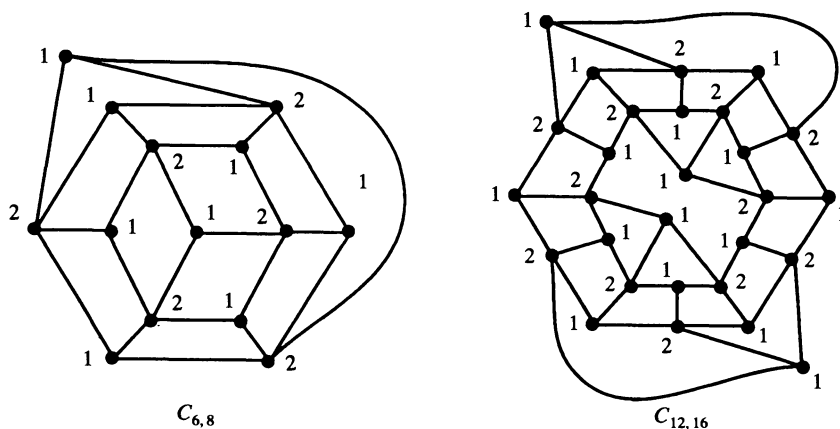


Figure 3

2.13 of [4].) There are 13 Archimedean solids, 11 of which can be obtained from the five Platonic solids by truncation. Archimedes' account of the 13 solids which bear his name is lost, presumably in the great fire of Alexandria. Heron writes that Archimedes ascribed the cuboctahedron to Plato. (See the historical notes in Chapter 2 of [2] and the Foreword of [6].) Kepler's book *The six-cornered snowflake*, published in 1609, includes what is believed to be the first complete list of the 13 Archimedean solids, giving them the names by which they are still known.

What properties will the dual graph $(C_{6,8})^*$ have? (For an introduction to duality for planar graphs, see Chapter 15 of [7], for example.) Since every vertex in $C_{6,8}$ has degree 3 or 4, every region in $(C_{6,8})^*$ will be bounded by 3 or 4 edges. Further, since each region of $C_{6,8}$ has 4 bounding edges, every vertex of $(C_{6,8})^*$ will have degree 4. Finally, since $C_{6,8}$ is bipartite, every edge of $(C_{6,8})^*$ will separate a triangle from a quadrilateral, so each triangle is surrounded by 3 quadrilaterals and each quadrilateral is surrounded by 4 triangles. Drawing $(C_{6,8})^*$ on a sphere, we find $(C_{6,8})^*$ is isomorphic to the *cuboctahedron*, an Archimedean solid (see FIGURE 4). The faces of the cuboctahedron are two colorable (so that the triangles can be colored red and the squares can be blue) and it can be constructed

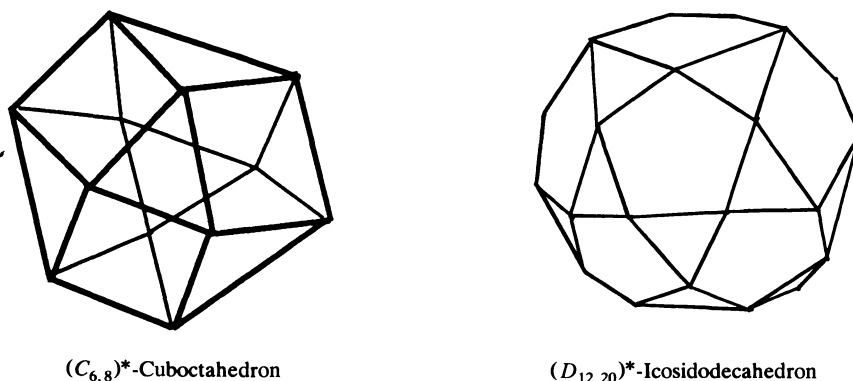


Figure 4

by truncating the eight vertices of an ordinary cube. More on building models of this solid (and many others) can be found in [4] or [6], both of which have very nice pictures.

We can also interpret $(C_{6k,8k})^*$ as the graphs of 3-dimensional solids for larger values of k . Again, each region will be bounded by either a triangle or a quadrilateral and each edge will separate a triangle from a quadrilateral. But two of the regions of $C_{6k,8k}$ are bounded by more than 4 edges, so two vertices of $(C_{6k,8k})^*$ will have degree larger than 4 (and so these solids will not be Archimedean). For example, in $(C_{12,16})^*$, there will be two vertices of degree 8 and 20 vertices of degree 4. The vertices of degree 8 (which correspond to the internal and external regions of $C_{12,16}$) will be incident to 4 triangles and 4 quadrilaterals each, while the remaining 20 vertices will have 2 triangles and 2 quadrilaterals surrounding them. The reader is encouraged to draw a picture (or even build a model) of $(C_{12,16})^*$, which is 'close to Archimedean.'

We can now apply the ideas developed above to the case where $a = 3$ and $b = 5$. Again, in the smallest possible example, the inequality in (§) will be replaced by equality. Thus, we are looking for a planar bipartite graph in which there are 60 edges, and so $m = 20$ and $n = 12$. From Euler's formula, this graph will have 30 regions, each of which is bounded by exactly 4 edges. We can construct such a graph using some of the ideas developed above; we denote the graph $D_{12,20}$. See FIGURE 5 for a drawing of this graph.

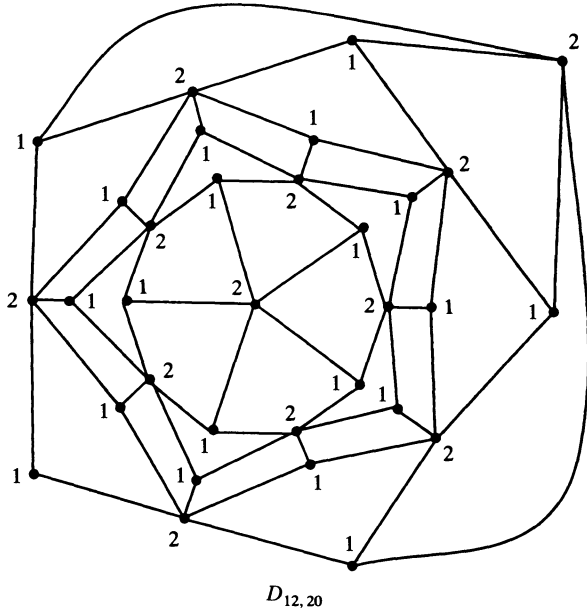


Figure 5

What does the dual graph look like this time? Proceeding as we did in examining $(C_{6,8})^*$, we note that every region in $(D_{12,20})^*$ will be bounded by 3 or 5 edges and every vertex of $(D_{12,20})^*$ will have degree 4. Further, every edge of $(D_{12,20})^*$ will separate a triangle from a pentagon (and so each triangle will be adjacent to 3 pentagons and each pentagon will be adjacent to five triangles).

Thus, $(D_{12,20})^*$ is also isomorphic to an Archimedean solid in which the regions (faces) are two-colorable (see FIGURE 4). This Archimedean solid is the *icosidodecahedron*, and it can be constructed by truncating the twenty vertices of a dodecahedron.

As in the $a = 3, b = 4$ case, we can generalize the procedure to construct an infinite family of planar bipartite graphs, denoted $D_{12k,20k}$, with $a = 3$ and $b = 5$. We modify $D_{12,20}$ in a way which is similar to the way we modified $C_{6,8}$ above to produce this family. See FIGURE 6 for a drawing of $D_{24,40}$, which has 64 vertices, 120 edges and 58 regions. (To simplify the picture, we have only labeled the 24 vertices of degree 5. All unlabeled vertices have degree 3 and should be labeled 1.) A model of the dual solid $(D_{24,40})^*$ will have 24 pentagons and 40 triangles. 56 of the 58 vertices of this solid will be incident to 2 triangles and 2 pentagons; the other 2 vertices will be incident to 4 triangles and 4 pentagons apiece. Again, this solid is 'almost' Archimedean in the same way $C_{12,16}$ is almost Archimedean.

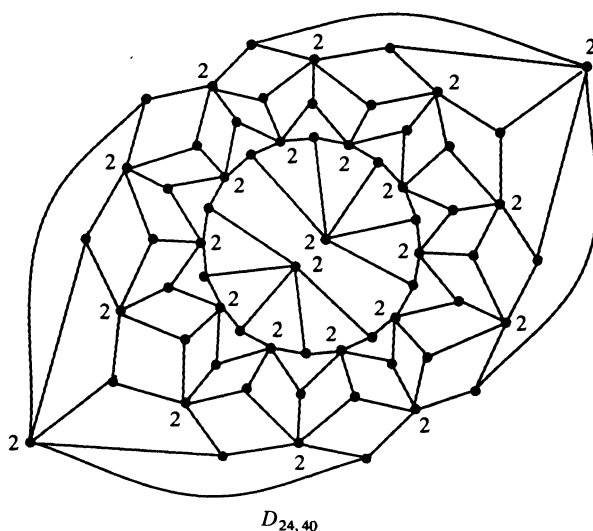


Figure 6

The two Archimedean solids used here, the cuboctahedron and icosidodecahedron, are *quasi-regular* in the sense that the two kinds of faces are all regular and a face of one kind is entirely surrounded by faces of the other kind. In fact, these two solids are the only convex quasi-regular solids. (This follows from our work above, or from Chapter 2 of [2].) We also remark that the graphs $C_{6,8}$ and $D_{12,20}$ are interesting enough to have their own names: $C_{6,8}$ is the *rhombic dodecahedron* and $D_{12,20}$ is the *triacontahedron*. Models of these two graphs (on spheres, considered as solids) can be constructed so that each face is a rhombus.

These two solids were discovered by Kepler around 1611; in fact, the rhombic dodecahedron occurs in nature as a garnet crystal. See Plate I in [2] for pictures of models of these solids. Coxeter [2] gives an elegant construction to create a model of the rhombic dodecahedron. Take two solid cubes and cut one of them into 6 square pyramids, based on the six faces, all sharing the center of the cube as the common apex of the pyramids. Now glue the bases of these 6 pyramids to the 6 faces of the other cube. The resulting solid is the rhombic dodecahedron. (See also

Table 4.12 of [4].) As far as I know, the models for the higher order graphs $C_{6k,8k}$ and $D_{12k,20k}$ and their duals (for $k > 1$) have not been considered before.

Finally we remark that the question concerning which of these graphs can actually be realized as gear arrangements, which was the motivation behind these examples, remains open. Many of the graphs in this paper can easily be shown to correspond to gear arrangements.

We conclude by summarizing the results of this section.

Proposition 5. *If G is a (connected) semi-regular planar bipartite graph in which every red vertex has degree a and every blue vertex has degree b (with $a < b$), then the only possibilities for a and b are the following:*

- (a) $a = 1$ and $b > 1$,
- (b) $a = 2$ and $b > 2$,
- (c) $a = 3$ and $b = 4$,
- (d) $a = 3$ and $b = 5$.

Furthermore, each of these possibilities can be realized.

ACKNOWLEDGMENTS. I would like to thank Elizabeth McMahon for her very useful comments concerning this paper and Lorenzo Traldi for some useful discussions. I dedicate this paper to Rebecca and Hannah, whose inspiration to me is limitless.

REFERENCES

1. J. A. Bondy and U. S. R. Murty, *Graph Theory with Applications*, North Holland, New York, 1976.
2. H. S. M. Coxeter, *Regular Polytopes*, (Macmillan, 1963) third edition, Dover, New York, 1973.
3. N. Hartsfield and G. Ringel, *Pearls in Graph Theory: A Comprehensive Introduction*, Academic Press, San Diego, 1990.
4. A. Pugh, *Polyhedra: A Visual Approach*, University of California Press, Berkeley, CA, 1976.
5. R. J. Trudeau, *Dots and Lines*, Kent State University Press, 1976.
6. M. J. Wenninger, *Polyhedron Models*, Cambridge University Press, New York, 1970.
7. R. J. Wilson, *Introduction to Graph Theory*, (Oliver & Boyd, 1972), third edition, Longman Scientific & Technical, Essex, England, 1985.

*Dept. of Mathematics
Lafayette College
Easton, PA 18042-1781
gordong@lafayett.bitnet*

A *sine qua non* for making mathematics exciting to a pupil is for the teacher to be excited about it himself; if he is not, no amount of pedagogical training will make up for the defect."

—R. L. Wilder

On the Kummer Solutions of the Hypergeometric Equation

Reese T. Prosser

One of the oldest, and still one of the most interesting, applications of group theory arises in the study of the transformations of an ordinary differential equation. If we know that a given differential equation admits a group of transformations, then we know that the solution set must admit that same group of transformations, and we can deduce properties of all the solutions from the properties of any one of them.

A case in point is offered by the celebrated hypergeometric equation (See Eq. (1) below), whose solutions include many of the most interesting special functions of mathematical physics. In his book [3], Einar Hille notes that this equation has a venerable history associated with such names as Gauss, Euler, Riemann, and Kummer. The hypergeometric equation is in fact a prototype: every ordinary differential equation of second order with at most three regular singular points can be brought to the hypergeometric equation by means of suitably chosen changes of variable [5].

In 1836 Kummer published a set of 6 distinct solutions of the hypergeometric equation. These include the hypergeometric function of Gauss, and all of them could be expressed in terms of Gauss's function (See Table 1 below). A useful summary of their basic properties is found in [1, p. 105ff.]. A glance at the list of these Kummer solutions reveals a rather complicated set of relationships which pleads for some simple explanation. We show here that the Kummer solutions are related by a finite group of transformations which serve to explain their relationships and to exemplify the use of transformation groups in the study of differential equations.

1. THE HYPERGEOMETRIC EQUATION. The hypergeometric equation has the form (See [1])

$$z(1-z)u''(z) + [c - (a+b+1)z]u'(z) - (ab)u(z) = 0. \quad (1)$$

Here a , b , and c are real or complex parameters which are independent of z . This equation has three regular singular points, at $z = 0$, 1 , and ∞ , so it belongs to the Fuchsian class. The roots of the corresponding indicial equations are:

$$0, 1 - c; \quad 0, c - a - b; \quad a, b.$$

Logarithmic singularities are to be expected at these singular points if c , $b - a$, or $c - a - b$ is an integer. We shall exclude these exceptional cases from our discussion here. It is known that every second order homogeneous linear differential equation with at most three singularities of regular type can be brought to this form by a suitable change of variable (See [5] for details).

$$\begin{aligned}
u_1 &= F(a, b; c; z) \\
T_1 u_1 &= (1 - z)^{-a} F(a, c - b; c; z/(z - 1)) \\
T_2 u_1 &= (1 - z)^{-b} F(c - a, b; c; z/(z - 1)) \\
T_3 u_1 &= (1 - z)^{c-a-b} F(c - a, c - b; c; z) \\
\\
u_2 &= F(a, b; a + b + 1 - c; 1 - z) \\
T_1 u_2 &= z^{-a} F(a, a + 1 - c; a + b + 1 - c; 1 - 1/z) \\
T_2 u_2 &= z^{-b} F(b + 1 - c, b; a + b + 1 - c; 1 - 1/z) \\
T_3 u_3 &= z^{1-c} F(b + 1 - c, a + 1 - c; a + b + 1 - c; 1 - z) \\
\\
u_3 &= z^{-a} F(a, a + 1 - c; a + 1 - b; 1/z) \\
T_1 u_3 &= z^{-a} (1 - 1/z)^{-a} F(a, c - b; a + 1 - b; 1/(1 - z)) \\
T_2 u_3 &= z^{-a} (1 - 1/z)^{c-a-1} F(1 - b, a + 1 - c; a + 1 - b; 1/(1 - z)) \\
T_3 u_3 &= z^{-a} (1 - 1/z)^{c-a-b} F(1 - b, c - b; a + 1 - b; 1/z) \\
\\
u_4 &= z^{-b} F(b + 1 - c, b; b + 1 - a; 1/z) \\
T_1 u_4 &= z^{-b} (1 - 1/z)^{c-b-1} F(b + 1 - c, 1 - a; b + 1 - a; 1/(1 - z)) \\
T_2 u_4 &= z^{-b} (1 - 1/z)^{-b} F(c - a, b; b + 1 - a; 1/(1 - z)) \\
T_3 u_4 &= z^{-b} (1 - 1/z)^{c-a-b} F(c - a, 1 - a; b + 1 - a; 1/z) \\
\\
u_5 &= z^{1-c} F(b + 1 - c, a + 1 - c; 2 - c; z) \\
T_1 u_5 &= z^{1-c} (1 - z)^{c-b-1} F(b + 1 - c, 1 - a; 2 - c; z/(z - 1)) \\
T_2 u_5 &= z^{1-c} (1 - z)^{c-a-1} F(1 - b, a + 1 - c; 2 - c; z/(z - 1)) \\
T_3 u_5 &= z^{1-c} (1 - z)^{c-a-b} F(1 - b, 1 - a; 2 - c; z) \\
\\
u_6 &= (1 - z)^{c-a-b} F(c - a, c - b; c + 1 - a - b; 1 - z) \\
T_1 u_6 &= z^{a-c} (1 - z)^{c-a-b} F(c - a, 1 - a; c + 1 - a - b; 1 - 1/z) \\
T_2 u_6 &= z^{b-c} (1 - z)^{c-a-b} F(1 - b, c - b; c + 1 - a - b; 1 - 1/z) \\
T_3 u_6 &= z^{1-c} (1 - z)^{c-a-b} F(1 - b, 1 - a; c + 1 - a - b; 1 - z)
\end{aligned}$$

TABLE 1. The Kummer Solutions.

Gauss in his Göttingen thesis (1812) showed that the function

$$\begin{aligned}
{}_2F_1(z) &= F(a, b; c; z) \\
&= \sum_{m=0}^{\infty} \frac{(a)_m (b)_m}{(c)_m (1)_m} z^m
\end{aligned} \tag{2}$$

is a solution of Eq. (1) which is regular at the origin. Here $(a)_m$ denotes the form

$$(a)_m = a(a + 1) \cdots (a + m - 1) = \frac{\Gamma(a + m)}{\Gamma(a)}, \tag{3}$$

and the series converges for $|z| < 1$. Note the symmetry:

$$F(a, b; c; z) = F(b, a; c; z).$$

In 1748, Euler had already shown that the function

$$u_1(z) = \int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tz)^a} dt \quad (4)$$

is also a solution of Eq. (1), provided that

$$0 < \operatorname{Re}(b) < \operatorname{Re}(c). \quad (5)$$

This solution is regular at the origin, and hence is a multiple of Gauss's solution ${}_2F_1(z)$. Evaluating these functions at $z = 1$, we see that

$$u_1(z) = B(b, c-b) {}_2F_1(z), \quad (6)$$

where $B(x, y)$ is the Beta function

$$\begin{aligned} B(x, y) &= \int_0^1 t^{x-1}(1-t)^{y-1} dt \\ &= \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}. \end{aligned} \quad (7)$$

We can verify Eq. (6) directly by expanding Eq. (4) in powers of z and integrating term by term.

If we make a simple change of variable in Eq. (4), of the form

$$\begin{aligned} t &\rightarrow 1-t, \\ t &\rightarrow t/(1-z-tz), \text{ or} \\ t &\rightarrow (1-t)/(1-tz), \end{aligned}$$

respectively, the resulting integral can be written again in the form of Eq. (4), and in this way we arrive at the *Euler transformations*:

$$\begin{aligned} F(a, b; c; z) &= (1-z)^{-a} F(a, c-b; c; z/(z-1)) \\ F(a, b; c; z) &= (1-z)^{-b} F(c-a, b; c; z/(z-1)) \\ F(a, b; c; z) &= (1-z)^{c-a-b} F(c-a, c-b; c; z). \end{aligned} \quad (8)$$

These transformations tell us that the solution $u_1(z)$ can be written in any of four equivalent forms.

2. KUMMER SOLUTIONS. In 1836 Kummer observed that any function of the form

$$u(z) = \int_{\Gamma} I(s, z) ds, \quad (9)$$

where

$$I(s, z) = \frac{s^{a-c}(s-1)^{c-b-1}}{(s-z)^a} \quad (10)$$

is a solution of Eq. (1) provided that the integrand is integrable along a simple path Γ joining any two of its singularities. This will be the case provided that

$$0 < \operatorname{Re}(b) < \operatorname{Re}(c) < \operatorname{Re}(a+1) < 2. \quad (11)$$

If we assume these conditions for the moment, then there are in general singularities of the integrand only at the points $s = 0, 1, \infty$, and z , and hence there are only $\binom{4}{2} = 6$ possible paths:

$$\begin{aligned} u_1(z) &= \int_1^\infty I(s, z) ds, & u_4(z) &= \int_z^\infty I(s, z) ds, \\ u_2(z) &= \int_{-\infty}^0 I(s, z) ds, & u_5(z) &= \int_0^z I(s, z) ds, \\ u_3(z) &= \int_0^1 I(s, z) ds, & u_6(z) &= \int_1^z I(s, z) ds. \end{aligned} \quad (12)$$

Each of these integrals reduces to the Euler form (Eq. (4)) by a suitable change of variable: By setting s equal, respectively, to

$$\begin{aligned} s &= 1/t, & s &= z/t, \\ s &= 1 - 1/t, & s &= zt, \\ s &= t, & s &= 1 - t(1 - z), \end{aligned}$$

we may write these six solutions, apart from inessential factors of the form $(-1)^\alpha$, in terms of Gauss's function $F(a, b; c; z)$ as follows:

$$\begin{aligned} u_1(z) &= F(a, b; c; z) \\ u_2(z) &= F(a, b; a + b + 1 - c; 1 - z) \\ u_3(z) &= z^{-a} F(a, a + 1 - c; a + 1 - b; 1/z) \\ u_4(z) &= z^{-b} F(b + 1 - c; b; b + 1 - a; 1/z) \\ u_5(z) &= z^{1-c} F(b + 1 - c, a + 1 - c; 2 - c; z) \\ u_6(z) &= (1 - z)^{c-a-b} F(c - a, c - b; c + 1 - a - b; 1 - z). \end{aligned} \quad (13)$$

These are Kummer's six distinct solutions of the hypergeometric equation. Each of them has four forms, which are related by the Euler transformations of Eq. (8), giving 24 forms in all (See Table 1 on page 536).

3. KUMMER TRANSFORMATIONS. Anyone looking for properties of the solutions of the hypergeometric equation and coming upon this table of the Kummer solutions for the first time might well be puzzled by the bewildering profusion it offers, and might well wonder if there were some way to organize this table so that the relations among the entries become clear.

One way to bring order out of chaos here is to introduce a finite group Γ_6 (the *Kummer group*) of six linear fractional transformations of the complex z -plane which transform the arguments of the Kummer solutions into one another. This group is generated by the elements

$$\begin{aligned} A(z) &= 1 - z \\ B(z) &= 1/z. \end{aligned} \quad (14)$$

If we represent the transformation $C(z) = (az + b)/(cz + d)$ as usual by the matrix

$$C = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then

$$A = \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$B = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}. \tag{15}$$

Note that $A, B \in GL(2, \mathbb{Z})$, $A^2 = B^2 = I$, and $\det A = \det B = -1$. The derivative $C'(z)$ of $C(z) = (az + b)/(cz + d)$ is

$$C'(z) = \det C / (cz + d)^2,$$

so that $\det C \cdot C'(z)$ is a square, whose root we denote by $W_C(z)$:

$$W_C(z) = \pm (\det C \cdot C'(z))^{1/2}, \tag{16}$$

where the sign is chosen so that $W_C(z) > 0$ when $0 < z < 1$. In particular,

$$W_A(z) = 1,$$

$$W_B(z) = 1/z. \tag{17}$$

The elements A and B generate a six-element subgroup Γ_6 of $GL(2, \mathbb{Z})$ consisting of

$$\Gamma_6 = \{I, A, B, AB, BA, ABA = BAB\}. \tag{18}$$

This group (the anharmonic group) can be visualized as the group of rotations of the regular hexagon (See below). The group Γ_6 admits a fundamental domain in the full complex plane \mathbb{C} which we can describe as follows. We define the domains D_1, D_2, D_3 and their conjugates D'_1, D'_2, D'_3 by:

$$D_1 = \{z: |z| < 1 \text{ \& } |1 - z| > 1/2\}$$

$$D_2 = \{z: |z| > 1 \text{ \& } |1 - z| < 1\}$$

$$D_3 = \{z: |z| > 1 \text{ \& } R1(z) < 1/2\}$$

$$D'_1 = \{z: |1 - z| < 1 \text{ \& } R1(z) < 1/2\}$$

$$D'_2 = \{z: |1 - z| > 1 \text{ \& } R1(z) > 1/2\}$$

$$D'_3 = \{z: |z| < 1 \text{ \& } R1(z) > 1/2\}.$$

Then these domains are disjoint, and, together with their boundaries, they fill up the plane (See FIGURE 1).

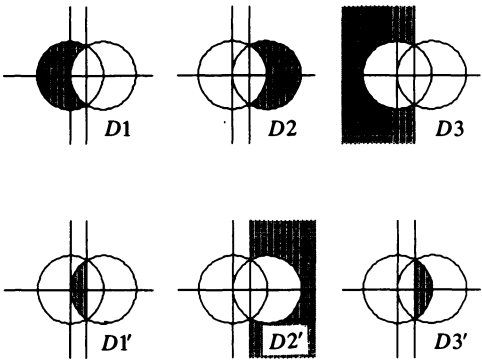
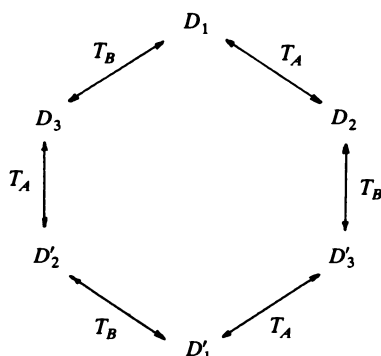


Figure 1. Kummer Domains

Moreover, they are related by the elements of the group Γ_6 :



(Note that $D'_i = T_{ABA}D_i$ for each value of i .)

At the same time we introduce a finite group of linear transformations of the parameter space \mathbf{C}^3 generated by the elements A and B of Γ_6 :

$$\begin{aligned} A(a, b, c) &= (a, b, a + b + 1 - c) \\ B(a, b, c) &= (a, a + 1 - c, a + 1 - b). \end{aligned} \quad (19)$$

If we denote by \mathbf{p} the vector $(a, b, c, 1)$ in \mathbf{C}^4 , then we can represent A and B by matrices:

$$\begin{aligned} A\mathbf{p} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ 1 \end{pmatrix} \\ B\mathbf{p} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ 1 \end{pmatrix} \end{aligned} \quad (20)$$

Note that the entries a and 1 in \mathbf{p} are fixed under A and B , and that if the entries of \mathbf{p} satisfy the conditions of Eq. (5), then so do those of $A\mathbf{p}$ and $B\mathbf{p}$. A modest calculation shows that A and B generate a six-element subgroup of $GL(4, \mathbf{Z})$ isomorphic to Γ_6 , as is already implicit in our notation.

Using these matrices, we can now define the group Γ_6 of transformations (the *Kummer transformations*) of the Kummer solutions in Eq. (13). If $G(a, b; c; z)$ is any function of z with parameters a, b , and c , not necessarily a solution of the hypergeometric equation, then we define for any $C \in \Gamma_6$ the transformation T_C by the formula

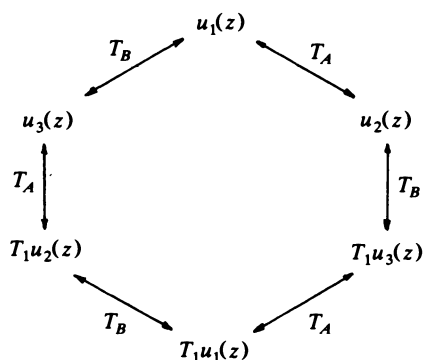
$$T_C G(\mathbf{p}, z) = W_C(z)^{p_2} G(C\mathbf{p}, C(z)), \quad (21)$$

where $p_1 = a$ is the first entry in \mathbf{p} , and the weight function $W_C(z)$ is given by

Eq. (16). With this definition it is straightforward to verify that

$$\begin{aligned}
 T_A u_1(z) &= T_A F(\mathbf{p}; z) \\
 &= F(a, b; a + b + 1 - c; 1 - z) \\
 &= u_2(z), \\
 T_B u_1(z) &= T_B F(\mathbf{p}; z) \\
 &= z^{-a} F(a, a + 1 - c; a + 1 - b; 1/z) \\
 &= u_3(z),
 \end{aligned} \tag{22}$$

More generally, one can verify directly that the following diagram is commutative:



To complete our analysis of the symmetries of the Kummer solutions, we introduce the transformation T_J , which realizes the symmetry $F(a, b; c; z) = F(b, a; c; z)$ of Eq. (2). Here

$$\begin{aligned}
 J(z) &= z, \\
 J\mathbf{p} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ a \\ b \\ c \end{pmatrix},
 \end{aligned} \tag{24}$$

$$W_J(z) = 1.$$

Then $J^2 = 1$ and $\det J = -1$, just as for A and B , and

$$T_J F(a, b; c; z) = F(b, a; c; z).$$

Then we can verify that

$$\begin{aligned}
 T_2 &= T_J T_1 T_J, \\
 T_3 &= T_1 T_2 = T_2 T_1,
 \end{aligned} \tag{25}$$

where T_2 and T_3 realize the second and third of the Euler transformations in Eq. (8); and that

$$\begin{aligned}
 T_J T_1 T_J &= T_2, \\
 T_J T_2 T_J &= T_1, \\
 T_J T_3 T_J &= T_3.
 \end{aligned} \tag{26}$$

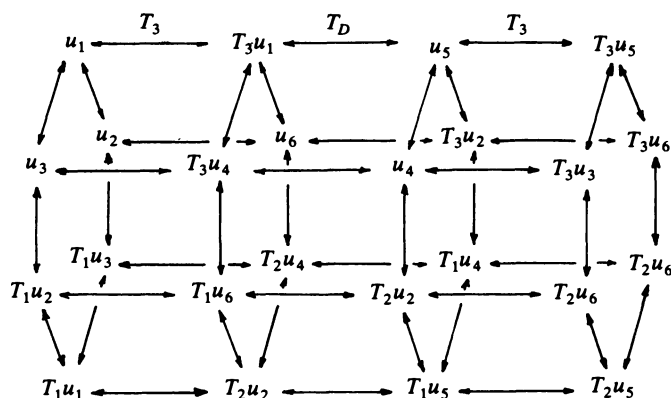
Finally, we can verify that

$$T_D = T_A T_3 T_A, \quad (27)$$

where the transformation T_D interchanges the independent solutions:

$$\begin{aligned} T_D u_1 &= u_5, \\ T_D u_2 &= u_6, \\ T_D u_3 &= u_4. \end{aligned} \quad (28)$$

We denote by Γ_{48} the 48-element subgroup of $GL(4, \mathbb{Z})$ generated by A , B , and J . We can now verify that the elements T_A , T_B , T_D , and T_J , as well as the elements T_1 , T_2 , and T_3 , all have order 2, and that the following master diagram is commutative.



In all, there are 8 orbits of the 6-element group Γ_6 , making a total of 48 possible forms for the Kummer solutions, consisting of those listed in Table 1 and their images under T_J . The transformation rule for any element C in the group Γ_{48} is given by Eq. (21).

Note added in proof. The referee, who is a much better historian than I, contributed the following remarks. "Euler not only knew that Eq. (4) is a solution of Eq. (1), he knew that the series Eq. (2) is a different form of this solution. Also Gauss claimed that Eq. (2) is a solution to an equation equivalent to Eq. (1) in his published paper from 1812 to 1813, but it was only in his unpublished work that he explicitly states the differential equation Eq. (1). He wrote it in different form. The most complete treatment of this equation to appear in print before Kummer's paper was in Pfaff's 1797 book. Euler found the third transformation in Eq. (8), in a posthumous paper which Pfaff edited (1797), and possibly in earlier work, although I have not found it earlier, but I have not found the first two transformations in Eq. (8) in work of Euler. A special case of them was given by Stirling in his 1730 book."

REFERENCES

1. A. Erdelyi et al., *Higher Transcendental Functions*, vol. 1, Chapt. II, The Bateman Manuscript Project, McGraw-Hill, New York, 1953.
2. H. Exton, *Multiple Hypergeometric Functions and Applications*, Chapt. 1, Wiley, New York, 1976.

3. E. Hille, *Ordinary Differential Equations in the Complex Domain*, Chapt. 6, Wiley, New York, 1976.
4. E. E. Kummer, Ueber die Hypergeometrisches Reihe, *J. f. Math.*, v. 15 (1836), pp. 39–83, 127–172.
5. E. G. C. Poole, *Introduction to the Theory of Linear Differential Equations*, Clarendon Press, Oxford, 1936.

*Department of Mathematics & Computer Science
Dartmouth College
Hanover, NH 03755*

From Richard Feynman

The following is a letter I received from Richard Feynman while I was a high school student at La Mirada High School in 1962. I was a Sophomore then, and Dr. Feynman had been invited to give a talk to our Math Club. The title of his talk that day was “What is one-half factorial?”.

I had no idea who Richard Feynman was, except that he worked at Cal Tech. Thinking at that time that I wanted to be a physicist, I wrote him a letter the next week asking about what colleges to attend and what mathematics classes to take. He kindly answered this letter.

September 26, 1962

Dear Mr. Farnum:

This is to answer your letter requesting information on colleges having good departments in physics. I am very poor at giving such advice, but I will try. It is best not to go to the same school for both undergraduate and graduate work, to develop a wider range of points of view. The schools that I can think of which have excellent physics graduate schools are Massachusetts Institute of Technology, California Institute of Technology, Columbia, Harvard, Princeton, Cornell, University of California at Berkeley, University of Chicago (and others which may not have come to mind). Don't worry about which is “best”—I don't know—they are all very good and each will give you a complete opportunity to learn and do as much as you can. For undergraduate work there is the above list of course, plus very many others, too numerous to mention. (If you have in mind any one in particular, not on the list above, let me know and I will give you my opinion of it.)

What mathematics to study? Whatever interests you the most at the moment is the best rule. Perhaps you will become fascinated, say, with number theory or topology, neither of which is used at all in physics. Then you will become a mathematician. Should we recommend that you don't study number theory or topology simply because it is not needed for preparation for theoretical physics? It is not necessary now to decide too closely your specialty. For your information, however, the mathematics of use in physics is the calculus and its extensions (as in the book “Advanced Calculus” by Woods) and later on vector analysis, partial differential equations, and other branches of analysis described in books with titles like “Mathematical methods of physics”. However, the emphasis should be somewhat more on how to do the mathematics quickly and easily, and what formulas are true, rather than the mathematicians' interest in methods of rigorous proof.

Sincerely,
Richard P. Feynman

*Submitted by Nicholas R. Farnum
Department of Management Science
California State University, Fullerton
Fullerton, CA 92634*

Reflections on a Mira

John W. Emert, Kay I. Meeks, and Roger B. Nelson

Reflective devices such as the Mira¹ are beginning to be used as a replacement for the compass and (unmarked) straightedge as tools for performing geometric constructions. These devices (see FIGURE 1) are generally constructed from a dark translucent plastic, producing an effect similar to that of looking into a car window on a bright, sunny day. When placed on a sketch, the device allows one to see both the figure and its reflection.

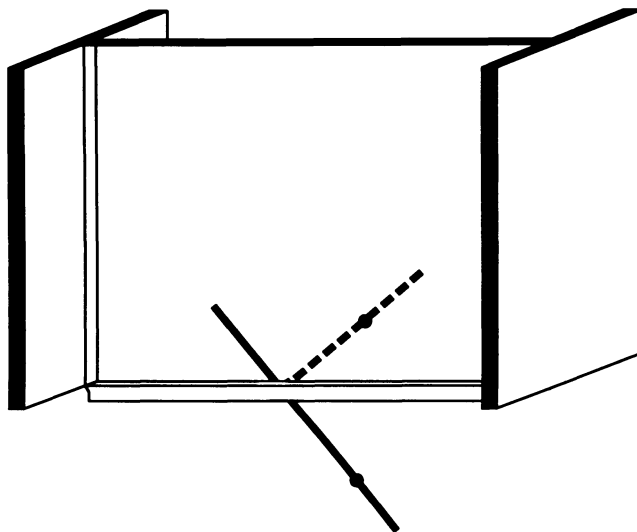


Figure 1

All isometries of the plane can be generated by reflections, and the Mira exploits this transformational approach more than the compass and straightedge. However, concerns over the mathematical basis of its replacement of the compass and straightedge are frequently expressed.

We address two issues related to these concerns. First, we formally describe the actions used in Mira constructions. Second, we algebraically characterize the set of Mira constructibles and note that the set of compass and straightedge constructible lengths is a proper subset of this set.

¹Mira is a registered trademark of the Mira-Math Company, Willowdale, Ontario, Canada. Any thin piece of colored, transparent hard plastic with a smooth finish is useable.

1. DECLARING PRIMITIVE ACTIONS. It is customary to assume that the primitive actions of the compass and straightedge are to draw the line through two given points and to draw the circle with given center and radius. When using the Mira, we identify the following as primitive actions:

- (1) Draw the line which reflects a given point to a given point.
- (2) Mark the reflection (i.e. mirror image) of a given point across a given line.
- (3) Draw a line passing through a given point which reflects a second given point to a given line, if such a line exists.
- (4) Draw a line which reflects two given points to respective given lines, if such a line exists.

Note that in Mira constructions, a circle is *known* by its center and radius, rather than by its graph. Likewise, a parabola is *known* by its focus and directrix. Therefore, primitive actions (3) and (4) above can be rephrased in the following manner.

- (3') Mark the intersections of a given line and the circle with given center and given radius.
- (4') Draw any lines which are simultaneously tangent to two parabolas, each determined by a given point and given line.

The following standard compass and straightedge constructions are among basic Mira constructions obtainable from the above primitive actions:

- (5) Draw the line through two given points.
- (6) Copy a segment onto a line, starting at a given point on that line.
- (7) Copy an angle to any given location and position.
- (8) Draw a line which reflects a given line to a given line.

For example, (5) can be obtained by drawing the diagonals of a rhombus having the given points as opposite vertices.

2. COMPARING CONSTRUCTIBLE RANGES. We will characterize those figures of classical Euclidean geometry which are constructible with a finite number of primitive actions, either using a compass and straightedge or using a Mira. We refer to such figures as being *C*-constructible or *M*-constructible, respectively. In either case, we note that a given angle α is constructible if and only if it is an angle of some triangle whose sides have constructible lengths. In fact, we may construct it as an angle of a right triangle whose legs have lengths $\cos \alpha$ and $\sin \alpha$. Since in either type of construction, segments can be transported (by (6)) and angles can be copied (by (7)), constructible figures may be formally described by characterizing constructible lengths, after designating a unit. We denote the set of *M*-constructible lengths by \mathcal{M} and the *C*-constructible lengths by \mathcal{C} .

Theorem 1. \mathcal{C} and \mathcal{M} are both subfields of the real numbers, \mathbb{R} .

Proof: One must only establish that if α and β are both constructible lengths then $\alpha - \beta$ and α/β are also constructible. The standard constructions [5, p. 428; 7, p. 400] involve only transporting lengths and copying angles and can therefore be carried out with either type of construction.

Recall that $F(\alpha)$ denotes the smallest subfield of a fixed algebraic closure of F containing both F and α . $F(\alpha)$ is a vector space over F with dimension equal to the degree of the unique monic irreducible polynomial over F of which α is a root. This is also called the degree of α over F and is denoted $[F(\alpha): F]$.

The classical result characterizing C -constructibles is [8, p. 213]:

Theorem 2. *A real number α is in \mathcal{C} if and only if there is a finite sequence of reals $r_1, r_2, \dots, r_n = \alpha$ for which*

$$[\mathbb{Q}(r_1, r_2, \dots, r_i) : \mathbb{Q}(r_1, r_2, \dots, r_{i-1})] = 2$$

for each integer i , $1 \leq i \leq n$.

Necessity follows from the observation that any newly C -constructed length results from the intersection of lines or circles determined by already C -constructed lengths and therefore is a root of some quadratic equation with these previously C -constructed coefficients. The algebraic condition is sufficient since the square root of any C -constructible length is also C -constructible. [5, p. 402]

The compass and straightedge construction of square roots which appears in [5] is easily modified into a Mira construction. Let $\alpha > 0$ be M -constructible. Construct OA with length α and locate P on the extension of \overrightarrow{AO} so that $|PO| = 1$. Let C be the midpoint of PA , and erect a line l perpendicular to PA through point O . Locate Q on l such that $|CQ| = |CA|$ (see FIGURE 2). Then angle PQA is a right angle since the triangle PQA is inscribed in the semicircle with radius $|CA|$ centered at C . It follows from the resulting similarity of triangles OPQ and OQA that $|OQ| = \sqrt{\alpha}$.

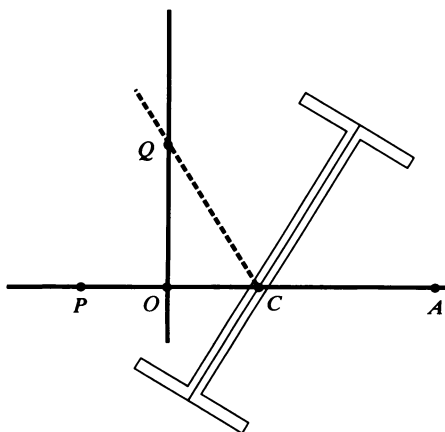


Figure 2

Since all roots of quadratics with already constructed coefficients are M -constructible it follows that all C -constructible lengths are M -constructible. That is:

Theorem 3. \mathcal{C} is a subfield of \mathcal{M} .

Note that the above theorem does not depend on use of the fourth primitive Mira action, since this action is not used in constructing quadratic roots. If the fourth action is not redundant, then the above containment must be proper. In fact, we will show in Sections 3 and 4 that \mathcal{M} consists of all α which result from a finite sequence of degree 2 or 3 extensions.

3. SOLVING ANY CUBIC EQUATION WITH A MIRA. We showed in Section 2 that the roots of any quadratic polynomial with Mira constructible coefficients can

be constructed by using combinations of Mira actions (1) through (3). With the inclusion of Mira action (4), we show in this section that the roots of any cubic polynomial can be found, a task not possible with compass and straightedge. We employ the method of Vieta (1591) (see [3]) as applied to the analysis of compass and marked ruler constructions in [9].

Any cubic equation can be transformed, by a rational transformation, to $x^3 + px + q = 0$, with discriminant $\delta = -4p^3 - 27q^2$. Recall that the sign of the discriminant indicates the order and number of real roots. When $\delta < 0$, the cubic has exactly one real root and can be transformed to a quadratic in w^3 by the substitution $x = w - p/3w$. When $\delta \geq 0$, all roots are real and the cubic can be reduced to the form $4z^3 - 3z - r = 0$, where $|r| \leq 1$. Expanding $r = \cos 3\alpha$ and substituting $z = \cos \alpha$ also leads to this form. Hence, the ability to construct cube roots and trisect angles allows the construction of real solutions to cubic equations with constructible coefficients.

The Mira action (4) provides a method to trisect an arbitrary angle. Such a construction can be found in [2], accompanied by a geometric justification. This construction can be modified to directly construct a root of $4z^3 - 3z - r$. It remains to show that the cube root of a constructible can be constructed.

Given the unit and a segment OY of length a , construct a perpendicular segment OX of unit length. Reflect Y across line OX to locate point Y' , and construct the line l_y through this point, parallel to OX . Similarly, reflect X across line OY to locate point X' , and construct the line l_x through this point, parallel to OY . Construct the Mira line which reflects X to l_x and Y to l_y , and mark the intersection of this line with OY as Z (see FIGURE 3). The segment OZ has length $\sqrt[3]{a}$.²

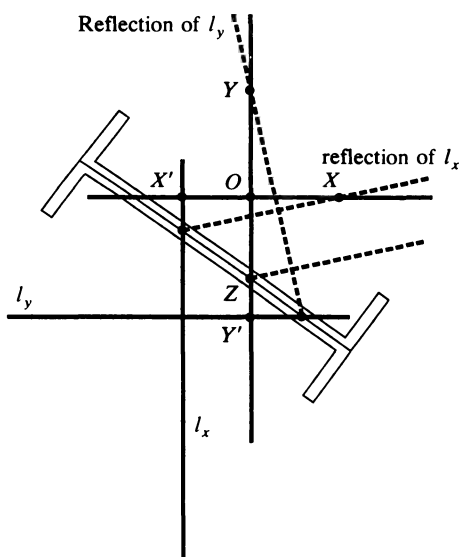


Figure 3

To justify this construction, first note that the pairs X, l_x and Y, l_y determine parabolas $x = y^2/4$ and $y = x^2/4a$, respectively. As these parabolas meet in two

²Thanks to M. H. Motto for his suggestions.

points, they have a unique common tangent. Comparing the general forms for the tangent line of each parabola, we obtain the common tangent line $y = -a^{-1/3}x - a^{1/3}$. This line meets the parabola $x = y^2/4$ at the point $(a^{2/3}, -2a^{1/3})$ and the parabola $y = x^2/4a$ at the point $(-2a^{2/3}, a^{1/3})$. As a tangent line common to two parabolas, its constructibility is guaranteed by (4'). This relationship between conics and the extraction of cube roots has ancient roots. The study of conic sections by the Greek mathematicians Hippocrates and Menæchmus arose from an attempt to duplicate the cube. See [1, pp. 106–109].

4. IDENTIFYING THE MIRA CONSTRUCTIBLE NUMBERS. We have just shown that a real number is M -constructible if it is a root of a quadratic or cubic equation with M -constructible coefficients. Are any other reals Mira constructible? No.

Since the first three primitive Mira actions can be accomplished with compass and straightedge, all M -constructibles involving only these actions are C -constructible. Therefore, if additional M -constructibles exist, they must result from applications of the fourth primitive action.

Recall that the fourth primitive action can be characterized as finding the common tangent to two *known* parabolas. If these parabolas are described by equations with coefficients in some subfield F of \mathbb{R} , we show that the coordinates of the points of tangency satisfy either cubic or quadratic equations with coefficients in F .

Any pair of parabolas with coefficients in a field F may be linearly transformed to another pair of parabolas whose central axes are either perpendicular or parallel. This linear transformation may be represented by a matrix with entries from F . $F(\alpha)$ and its image under such a transformation will have the same degree over F . Hence, we need to consider only those cases where the central axes are parallel or perpendicular.

When the central axes are perpendicular, we may assume that the axes intersect at the origin and express the parabolas as $y^2 = a(x - h)$ and $x^2 = b(y - k)$, where a, b, h and k are all in F . These parabolas meet the common tangent in points (x_1, y_1) and (x_2, y_2) , respectively. The common tangent has slope $a/2y_1 = 2x_2/b$. Computing the y -intercept in two ways yields the relationship $2k - y_2 = (a/2y_1)(x_1 - 2h)$. It follows from the resulting four equations in four unknowns that y_1 satisfies

$$8y^3 - 16ky^2 - 8ahy + a^2b = 0$$

and x_1 satisfies

$$64a(x - h)^3 - 128(ah + 2k^2)(x - h)^2 + 32(2h^2 + abk)(x - h) - a^2b^2 = 0.$$

Coordinates x_2 and y_2 satisfy analogous cubics having coefficients in F . A similar attack on the case when the central axes of the parabolas are parallel shows that all coordinates x_1, y_1, x_2, y_2 satisfy quadratic equations.

We see, therefore, that there are no other M -constructible numbers. That is:

Theorem 4. *A real number α is in \mathcal{M} if and only if there is a finite sequence of reals $r_1, r_2, \dots, r_n = \alpha$ for which*

$$[\mathbb{Q}(r_1, r_2, \dots, r_i) : \mathbb{Q}(r_1, r_2, \dots, r_{i-1})] = 2 \text{ or } 3$$

for each integer i , $1 \leq i \leq n$.

Corollary. *A regular n -gon is Mira-constructible if and only if n has the factorization $2^s 3^t p_1 \dots p_k$, where $p_1 \dots p_k$ are distinct primes of the form $2^u 3^v + 1$.*

From the classical result for \mathcal{C} (Theorem 2), Gauss showed that a regular n -gon is constructible with compass and straightedge if and only if n has the factorization $2^s p_1 \dots p_k$, where $p_1 \dots p_k$ are distinct primes of the form $2^u + 1$ (Fermat primes). [8, p. 265]. Gleason recently showed that a regular n -gon is constructible with compass, straightedge, and angle-trisector if and only if n has the factorization as stated in the Corollary above [6].

The construction of a regular n -gon corresponds to adjoining a primitive n -th root of unity, resulting in a *normal* extension. Gleason shows that any normal extension of degree 3 of a constructible field is constructible by use of straightedge, compass, and angle-trisector, by observing that the corresponding irreducible polynomial has all real roots. Since elementary Mira constructions correspond to extensions (not necessarily normal) of degree 2 or 3 only, the Mira-constructible regular polygons are precisely those constructible by straightedge, compass, and angle trisector.

CONCLUSION. We have shown that constructions performed with reflective devices such as the Mira can be formally defined, analyzed and justified mathematically. In addition, the set of Mira constructibles extends the set of compass and straightedge constructibles in the same manner that the compass and straightedge constructibles extend the rationals.

REFERENCES

1. C. B. Boyer and U. C. Merzbach, *A History of Mathematics*, 2nd ed., Wiley, New York, 1989.
2. I. M. Dayoub and J. W. Lott, *Geometry: Constructions and Transformations*, Dale Seymour, Palo Alto, 1977.
3. L. E. Dickson, *New First Course in the Theory of Equations*, Wiley, New York, 1939.
4. H. W. Eves, *A Survey of Geometry*, Allyn and Bacon, Boston, 1972.
5. J. B. Fraleigh, *A First Course in Abstract Algebra*, Addison Wesley, Reading, Massachusetts, 1989.
6. A. M. Gleason, *Angle Trisection, the Heptagon, and the Triskaidecagon*, *The American Mathematical Monthly* **95** (1988), 185–194.
7. T. W. Hungerford, *Abstract Algebra*, Saunders, Philadelphia, 1990.
8. N. Jacobson, *Basic Algebra*, W. H. Freeman, San Francisco, 1974.
9. R. C. Yates, *Geometrical Tools*, Educational Publishers, St. Louis, 1949.

Department of Mathematical Sciences

Ball State University

Muncie, IN 47306-0490

Emert: 00jwemert@leo.bsuvc.bsu.edu

Meeks: 00kimeeks@leo.bsuvc.bsu.edu

Nelson: 00rbnelson@leo.bsuvc.bsu.edu

Buffon Noodles

Ed Waymire

1. INTRODUCTION. During the 1992 spring term, physicist Corinne Manogue gave a colloquium talk to our mathematics department in which she surveyed various mathematical notions which arise in a string theorist's approach to quantum field theory. All this talk of strings and probability by Corinne made an impression on my colleague Robby Robson, who had an interesting hallway question the following day. Robby was wondering if I knew what would happen if Buffon had tossed a noodle in place of a needle (actually Buffon tossed baguettes)? Here the image of noodles is that of the long thin stringy type which always seem an entangled mess. So this story will feature entangled noodles.

2. BUFFON NEEDLE. Let's recall the Buffon needle problem. Snell ([7]) has a very nice treatment together with original references and a wonderful dose of historical perspective blended with modern computer simulations; that's where I learned about the baguette. One tosses a needle of length L onto a plane surface (floor) marked by parallel lines of width $D > L$ units apart. One asks for the probability of the event C that the needle intersects its closest line. The answer is a (linear) function

$$p_0(x) := P(C) = \frac{2L}{\pi D} = \gamma_0 x \quad (1)$$

of the ratio $x = L/D$ with slope $\gamma_0 = 2/\pi$ ($\approx .636$). A particularly interesting aspect of this example is that one obtains a statistical method of estimating numerical values of π from the crossing frequency of repeated needle tossings.

One needs a model of the needle tossing experiment to compute the probability in (1). For this, let Y denote the vertical distance from the midpoint of the needle to the nearest parallel and let Θ denote the acute angle made by the needle and the nearest parallel. Then one assumes that (Y, Θ) is uniformly distributed over the range of values $[0, D/2] \times [0, \pi/2]$. Now the problem is a geometry problem. It is interesting for the ensuing discussion that the crossing probability (1) does not change if one assumes that an *endpoint* of the needle is randomly (uniformly) selected on a fixed vertical line and, independently, the angle made by the needle with the vertical line is obtained by a random rotation from 0 to 2π .

3. BROWNIAN NOODLE. While others (eg., see [1], [5]) have considered "noodle" extensions defined by bending needles into various specific convex shapes to be tossed, we shall allow the noodles to become randomly *tangled*. To begin, let us first consider what happens in the case of a two-dimensional standard Brownian motion $\{B(t) = (B_1(t), B_2(t))\}$ over a time interval of unit length and started at

random along a vertical line of length D , FIGURE 1a. Although the total length of the Brownian noodle is ∞ , one may introduce L as a scale parameter by taking

$$S(t) := \frac{\sqrt{2} L}{2} B(t), \quad 0 \leq t \leq 1. \tag{2}$$

The answer in this case is

$$p_1(x) := P(C) = \frac{8}{\pi^2} \sum_{m=1}^{\infty} \frac{1}{(2m-1)^2} (1 - e^{-(2m-1)^2(\pi^2/4)(L/D)^2}). \tag{3}$$

Now the parameter L is a scale parameter and the gap is D . The crossing probability in (3) depends only on the ratio L/D . The extra factor $\sqrt{2}/2$ is being introduced here for the convenience of making comparisons with computations later. For convenience now let $L' := \sqrt{2} L/2$.

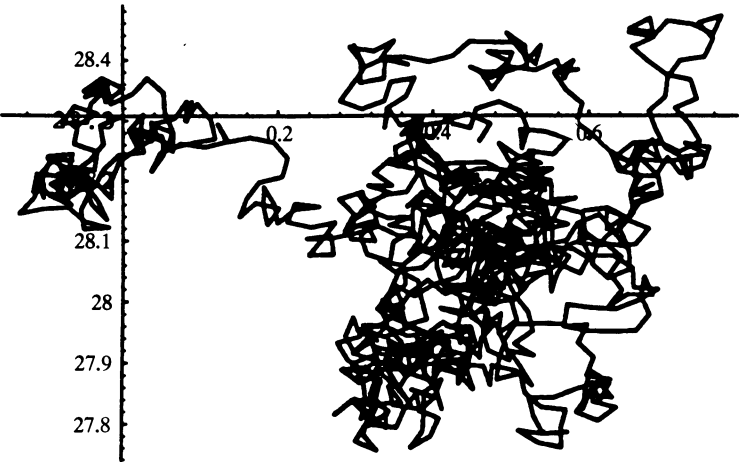


Figure 1a. Sample Realization of Brownian Noodle Toss.

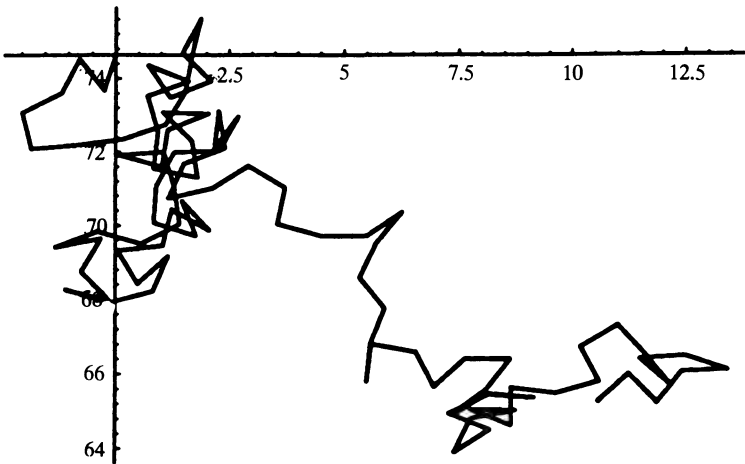


Figure 1b. Sample Realization of 100 Strand Noodle Toss.

To obtain the solution (3) one first observes that it is enough to consider the one-dimensional Brownian motion $\{B_1(t)\}$ on the interval $[0, D/L']$ with uniform initial distribution on $[0, D/L']$. We will need the joint distribution of (M_1, m_1) where

$$M_1 = \max_{0 \leq t \leq 1} B_1(t), \quad m_1 = \min_{0 \leq t \leq 1} B_1(t). \quad (4)$$

For then the answer is simply $1 - P(0 < m_1 < M_1 < D/L')$. There are various ways to compute the joint distribution of the maximum and minimum in (4). One approach is to first compute the transition probabilities of the Brownian motion on $[0, D/L']$ with absorbing boundaries at 0, D/L' and starting at $0 < x < D/L'$. Since these transition probabilities satisfy the heat equation on the interior of the interval with Dirichlet boundary conditions one may compute the following eigenfunction expansion given in ([2], p. 412–413):

$$p(t; x, y) = 2 \frac{L'}{D} \sum_{m=1}^{\infty} \exp\left\{-\frac{m^2 \pi^2 t}{2} \left(\frac{L'}{D}\right)^2\right\} \sin\left(\frac{m \pi L' x}{D}\right) \sin\left(\frac{m \pi L' y}{D}\right), \quad (5)$$

$$\left(0 < x, y < \frac{D}{L'}\right).$$

Now the desired probability (3) is the complementary probability to that obtained by performing the indicated integrations in

$$P\left(0 < m_1 < M_1 < \frac{D}{L'}\right) = \frac{L'}{D} \int_0^{D/L'} \int_0^{D/L'} p(1; x, y) dy dx. \quad (6)$$

Here one also uses the familiar fact that the sum of reciprocal squares is $\pi^2/6$, so that restricting the sum to reciprocals of squares of odd numbers yields

$$\frac{\pi^2}{6} - \frac{1}{4} \frac{\pi^2}{6} = \frac{\pi^2}{8}.$$

FIGURE 2a contains a plot of the two probabilities (1) and (3) as a function of the ratio $x = L/D$. The corresponding result of a computer simulation is given in FIGURE 2b.

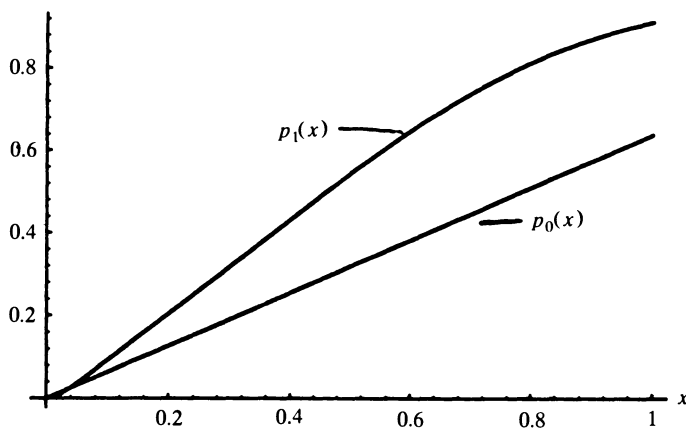


Figure 2a. Theoretical Brownian Noodle Crossing Probabilities $p_1(x)$.

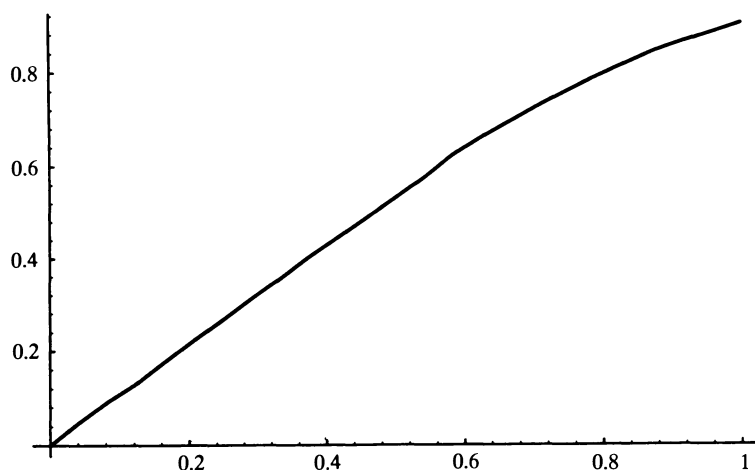


Figure 2b. 5000 Toss Simulated Crossing Frequencies for 100 Strand Brownian Noodle.

4. “REAL” NOODLES. Next let us consider a real string of n needles of lengths L/\sqrt{n} each, and strung together in independent random directions. The total length is then $n(L/\sqrt{n}) = L\sqrt{n}$. In order to preserve the ratio as $x = L/D$, the gap between parallel lines is chosen as $D\sqrt{n}$. Since we already know the answer for $n = 1$ by (1), let's do the problem for $n \gg 1$. In this case we first note that the *functional central limit theorem* provides an approximation to the distribution of the string of needles by the Brownian noodle. In particular, observe that the lengths of the projections of the needle are i.i.d. with, up to the scale factor L/\sqrt{n} , the arcsine distribution $P(Y \in dy) = (1/\pi\sqrt{1-y^2}) dy$. This distribution has mean 0 and variance $1/2$. Thus the Brownian motion approximation has diffusion coefficient $L^2/2$. This explains the scaling factor $(\sqrt{2}/2)L$ introduced earlier.

5. FIRST ROUND APPROXIMATIONS. In the limit the approximation involves a noodle $\{S(t): 0 \leq t \leq 1\}$ of length ∞ in a gap of width ∞ , but in the ratio $x = L/D$. This is how string theorists discussing quantum field theory can sound! One might “expect” the noodle crossing frequency to be estimated by $p_1(x/\sqrt{n})$ as this corresponds to formally replacing D by $\sqrt{n}D$ in (3). This curve is given in FIGURE 3a with $n = 100$. FIGURE 3b provides the crossing frequencies $\bar{p}_5(x)$ obtained from a simulation of 5000 tosses with $n = 100$ for comparison. However, since the effect of fixing the gap to length ratio in the above computation is a second rescaling of the string of noodles by $1/\sqrt{n}$, the central limit theorem approximation by the Brownian noodle is not justified. On the other hand, the formula (3) is applicable to strings composed of a large number n of needles of (fixed) lengths x each tossed onto a surface with gap width \sqrt{n} . This fact is illustrated by the simulation in FIGURE 2b of the corresponding crossing frequencies $\bar{p}_1(x)$ for a Brownian noodle with $n = 100$, $N = 5000$; i.e. these are the crossing frequencies of $-\sqrt{n}/2$ or $\sqrt{n}/2$ by a string of n randomly (uniformly) oriented needles of length x each with endpoint randomly distributed over $[-\sqrt{n}/2, \sqrt{n}/2]$.

Due to the additional rescaling of the gap width, there will be no loss in generality to consider the crossing probability of strings composed of n needles of length $x \leq 1$ each and a gap of width n . A quick and clean (rigorous) upper bound

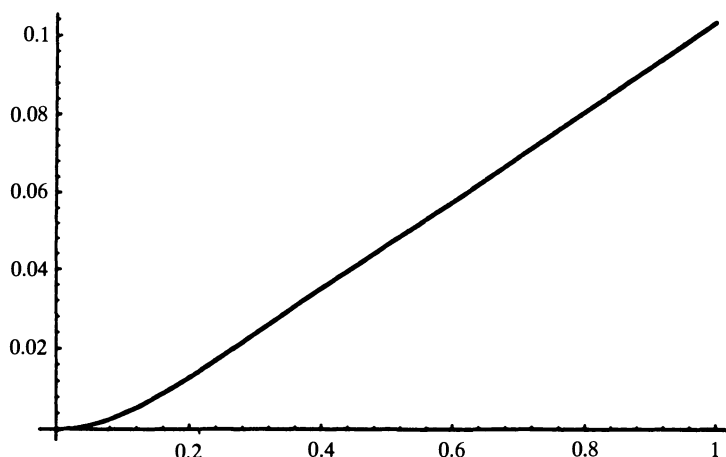


Figure 3a. Theoretical Brownian Noodle Crossing Probabilities $p_1(x/\sqrt{100})$.

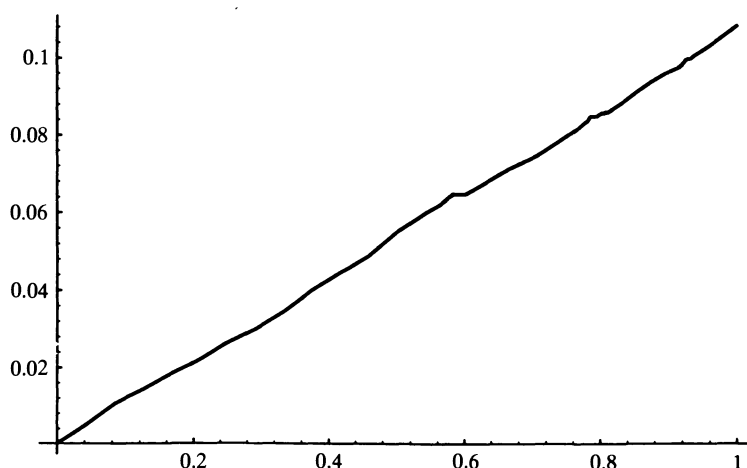


Figure 3b. 5000 Toss Simulated Crossing Frequencies for 100 Strand Noodle.

on the crossing probability can be obtained with *Doob's maximal inequality* ([2], pp. 49–52). In particular one obtains, using translation invariance of the problem to pose it with one end started uniformly in $[-n/2, n/2]$, that

$$P(C) \leq \frac{4}{n}x^2 + \frac{2}{3}. \quad (7)$$

For large n this bounds the probability by the essentially constant value $2/3 +$. Recall that the maximum crossing probability for the needle (1) is $2/\pi \approx 2/3 -$. To make this estimate, let $\{S_n\}$ denote the random walk starting at 0 with displacement distribution $P(Y \in dy) = (1/\pi\sqrt{1-y^2})dy$, and let $Z^{(n)} \equiv nU$ be uniformly distributed on $[-n/2, n/2]$ independently of the random walk. The details of the distribution of Y do not play a role beyond first and second moments. The essential observation is that by symmetry $\{xS_k + Z^{(n)}: k = 0, 1, \dots\}$ is a *martingale*. Thus, Doob's inequality gives (7), after conditioning on $Z^{(n)}$ and

using symmetry, as follows:

$$\begin{aligned}
 P(C) &= EP\left(\max_{k \leq n} (xS_k + Z^{(n)}) \geq \frac{n}{2}\right) \cup \left\{\min_{k \leq n} (xS_k + Z^{(n)}) \leq -\frac{n}{2}\right\} | Z^{(n)} \\
 &\leq EP\left(\max_{k \leq n} (xS_k + Z^{(n)}) \geq \frac{n}{2} | Z^{(n)}\right) + EP\left(\min_{k \leq n} (xS_k + Z^{(n)}) \leq -\frac{n}{2} | Z^{(n)}\right) \\
 &= 2EP\left(\max_{k \leq n} (xS_k + Z^{(n)}) \geq \frac{n}{2} | Z^{(n)}\right) \leq 2 \frac{4}{n^2} E(xS_n + Z^{(n)})^2 \\
 &= \frac{8}{n^2} (x^2 ES_n^2 + E(Z^{(n)})^2) = \frac{8}{n^2} \left(\frac{x^2 n}{2} + \frac{n^2}{12}\right) = \frac{4}{n} x^2 + \frac{2}{3}. \quad (8)
 \end{aligned}$$

A bound as universal as (7) will be much too large for many cases of interest. More precise estimates of the crossing probability upper bound are given in the next section.

For a lower bound, let $D(\beta)$ denote the event that the initial end of the string falls within βx , $\beta \leq 1$, units from the boundary. Then $D(\beta)$ has probability $2\beta x/n$. Also, from the geometry $P(C|D(\beta)) \geq (\cos^{-1}(\beta)/\pi)$ since, for any n , crossing is implied by the occurrence of the angle of the first needle between $\pm \cos^{-1}(\beta)$. Thus, for any $0 \leq \beta \leq 1$, $P(C) \geq 2\beta x/n(\cos^{-1}(\beta)/\pi)$ so that

$$P(C) \geq \lambda(n)x, \quad \text{where } \lambda(n) = \frac{2}{n\pi} \max_{0 \leq \beta \leq 1} \beta \cos^{-1}(\beta) \approx \frac{.357}{n}. \quad (9)$$

6. LARGE DEVIATIONS. As noted above, the functional central limit theorem approximation involved in replacing the string of needles by a Brownian motion for large n provides probabilities of events involving $O(1)$ fluctuations in the shape of the string of needles; the individual needle lengths already being $O(1/\sqrt{n})$. However, the crossing probabilities of $\sqrt{n}D$ involve $O(\sqrt{n})$ deviations from this picture which are not obtained by this approximation. These fluctuations have probabilities which also depend on the details of the angular distribution of the needles in the string. The precise probability computation desired here is a *large deviation problem* of the type which arises in the actuarial mathematics of insurance risk. In fact, the classic work of Cramer on large deviations was motivated by actuarial problems. Posed this way, the noodle problem also has an interesting twist in that it involves a “two-sided ruin event.”

To see how the distribution will depend on the angle distribution, consider the degenerate case in which the angles made by the needles comprising the string with the horizontal are each $\pi/2$ (nonrandom). Then the noodle is a needle with fixed orientation and the only randomness is in the distribution of the endpoint; see FIGURE 4a. For this case one readily finds that $p_3(x) = P(C) = x$. (This probability is unchanged if one randomizes the choice of endpoint). To see how large deviation estimates go in a simple but less trivial situation for which explicit computations are still possible, consider the case in which the needles comprising the string make angles $\pm\pi/4$ with the horizontal with equal probabilities; see FIGURE 4b for a sample outcome of a simulation of a single toss with $n = 100$. Since the variance in step size is one in this case, the rescaling by $\sqrt{2}/2$ is unnecessary for the comparison with the (nonrigorous) central limit theorem prediction; i.e. replace x by $x/2$ in FIGURE 3a to correct for the $\sqrt{2}/2$ before comparing with the observed FIGURE 5. In this case the lengths of the projections of the needle are i.i.d. with, up to the scale factor L , a Bernoulli distribution

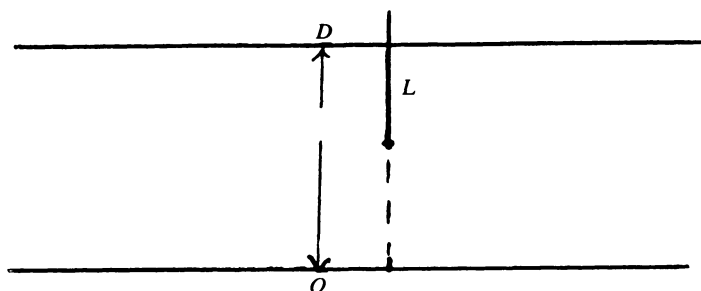


Figure 4a. Sample 0 Degree Single Strand Noodle Realization.

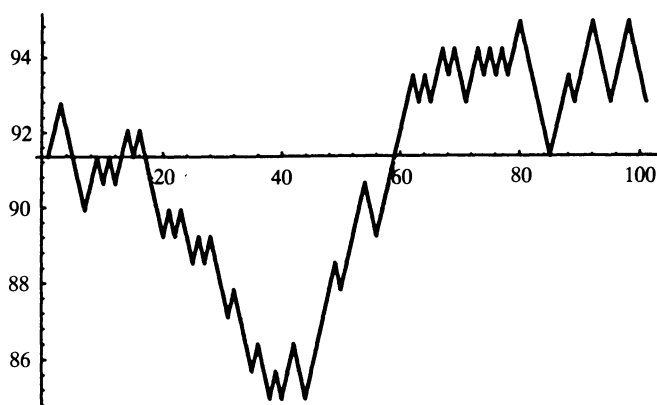


Figure 4b. Sample 100 Strand 45 Degree Noodle Realization.

$P(Y_1 = +1) = P(Y_1 = -1) = 1/2$. Let $S_n := (Y_1 + Y_2 + \cdots + Y_n)$, $S_0 = 0$. More generally, assume $\phi(s) = Ee^{sY_1}$ finite and note for i.i.d. $\{Y_n\}$, $(\phi(s))^n = Ee^{sS_n} \geq e^{nsb}P(S_n \geq nb)$. That is, for arbitrary s

$$P(S_n \geq nb) \leq e^{n(\log(\phi(s)) - sb)}. \quad (10)$$

In particular,

$$P(S_n > nb) \leq e^{n\chi^*(b)}, \quad \text{where} \quad \chi^*(b) = \inf_s (\log Ee^{sY} - sb). \quad (11)$$

The reverse (asymptotic) inequality for $b > EY_1$ yielding

$$\frac{1}{n} \log P(S_n > nb) \rightarrow \chi^*(b) \quad \text{as } n \rightarrow \infty, \quad (12)$$

is the often cited *Cramer-Chernoff theorem* on large deviations of S_n/n from EY_1 . That is to say, $P(S_n > nb) \sim e^{n\chi^*(b)}$ as $n \rightarrow \infty$; the rate $-\chi^*(b)$ is a *large deviation rate*. The reader may consult the developments given in ([3], pp. 147–148) and ([4], pp. 57–63) for some basics of large deviation theory.

Using the reflection principle for simple symmetric random walk one may also check that ([2], p. 10, 70), $P(\max_{k \leq n} S_k \geq b) = 2P(S_n \geq b) - P(S_n = b)$. In particular one obtains from this,

$$P\left(\max_{k \leq n} S_k \geq b\right) \leq 2P(S_n \geq b). \quad (13)$$

These results will serve as our basic tools. Notice that for large n it is no more than a venal sin to think of the inequality (13) as equality; i.e. the resulting probability bound may be expected to be close to the actual probability.

First let us compute the so-called *Legendre transform* $\chi^*(b)$ of the function $\chi(s) = \log Ee^{sY}$ required in (11). One may easily check by calculating limits as $s \rightarrow \pm\infty$ that $\chi^*(b) = -\infty$ if $|b| > 1$. For $|b| < 1$, equating the derivative with respect to s of $\{\log((e^s + e^{-s})/2) - sb\}$ to 0 and solving for s gives $e^{2s} = ((1 + b)/(1 - b))$ and therefore

$$\chi^*(b) = -\frac{1+b}{2}\log(1+b) - \frac{1-b}{2}\log(1-b). \quad (14)$$

As before let $Z^{(n)} = nU$ be uniformly distributed over $[-n/2, n/2]$ and independent of $\{S_n\}$. Then one has along the lines of (8) that

$$P(C) \leq 2P\left(\max_{k \leq n} S_k \geq n\left\{\frac{1}{2x} + \frac{U}{x}\right\}\right) \leq 4P\left(S_n \geq n\left\{\frac{1}{2x} + \frac{U}{x}\right\}\right), \quad (15)$$

first using the symmetry of the distribution of both $\{S_n\}$ and U , and then the reflection principle. By conditioning on U , using (15), the argument for (11) conditionally, and then (14), one has after taking expected values (with respect to U),

$$\begin{aligned} P(C) &\leq 4EP\left(S_n \geq n\left\{\frac{1}{2x} + \frac{U}{x}\right\} \middle| U\right) \leq 4Ee^{n\chi^*(1/2x + U/x)} = 4x \int_0^{1/x} e^{n\chi^*(y)} dy \\ &= 4x \int_0^1 \exp\left\{-\frac{n}{2}[(1+y)\log(1+y) + (1-y)\log(1-y)]\right\} dy := \gamma_4(n)x. \end{aligned} \quad (16)$$

Note that the (concave) function

$$\chi(y) = -(1/2)[(1+y)\log(1+y) + (1-y)\log(1-y)]$$

and its first derivative are zero at $y = 0$. The second derivative at 0 is -1 . As a result one may bound by a Taylor's approximation $-(1/2)y^2$ to obtain by a simple change of variable and the fact that

$$\begin{aligned} \int_0^{\sqrt{n}} e^{-(1/2)z^2} dz &\sim \sqrt{2\pi}/2, \\ \gamma_4(n) &= 4 \int_0^1 \exp\left\{-\frac{n}{2}[(1+y)\log(1+y) + (1-y)\log(1-y)]\right\} dy \\ &\leq 2\sqrt{2\pi} n^{-1/2} \approx 5.01 n^{-1/2}. \end{aligned} \quad (17)$$

Observe from this that the crossing probabilities are theoretically predicted to be smaller than the universal bound (7) for 45 Degree strings consisting of n strands. FIGURE 5 depicts the crossing frequencies $\bar{p}_4(x)$ for a simulation of $N = 1000$ tosses with $n = 100$. A sample realization of a toss was plotted in FIGURE 4b. Finally, observe that the argument used to arrive at the lower bound (9) can be simply adapted to this case to give

$$\frac{x}{n} \leq P(C) \leq \gamma_4(n)x \sim \frac{2\sqrt{2\pi}}{\sqrt{n}}x. \quad (18)$$

To extend these estimates to the case of uniform angles on $[0, 2\pi)$ first requires a general inequality of the form (15). Such an extension is possible for symmetric

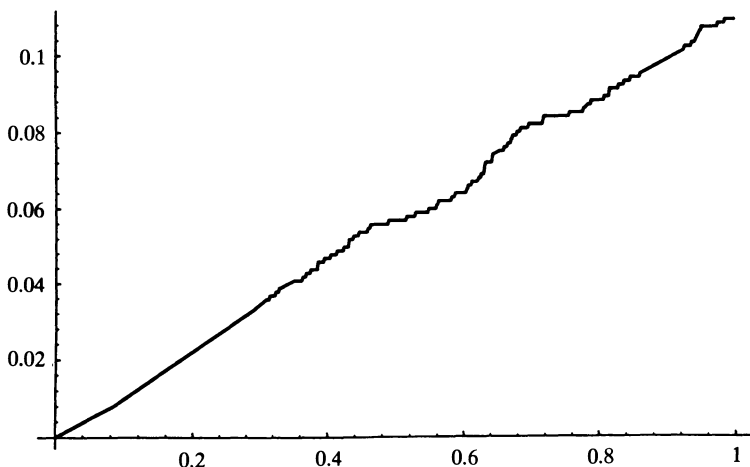


Figure 5. 5000 Toss Simulation Crossing Frequencies for 100 Strand 45 Deg Noodle.

random variables (*Levy's inequality*), ([5], p. 14; [2], Exc. #4, p. 70). The corresponding bound in place of (16) for this case is, following the first two inequalities of (16),

$$P(C) \leq 4Ee^{n\chi^*(1/2x+U/x)} = 4x \int_0^{1/x} e^{n\chi^*(y)} dy = 4x \int_0^1 e^{n\chi^*(y)} dy, \quad (19)$$

where now for $|y| > 1, \chi^*(y) = -\infty$, but for $|y| < 1$

$$\chi^*(y) = \inf_s \left\{ \log \left(\frac{1}{\pi} \int_{-1}^1 \frac{e^{st}}{\sqrt{1-t^2}} dt \right) - sy \right\}. \quad (20)$$

The crossing frequencies $\bar{p}_5(x)$ obtained from a simulation of 5000 tosses with $n = 100$ strands were provided earlier in FIGURE 3b. To obtain the asymptotic value of $\gamma_5(n)$ one may proceed as in the above case to compute the values of $\chi^*(y)$ and its first two derivatives at $y = 0$. This is most easily accomplished by the following Legendre transform *duality equations*: Let

$$\chi(s) = \frac{1}{\pi} \int_{-1}^1 \frac{e^{st}}{\sqrt{1-t^2}} dt = \frac{1}{\pi} \int_{-(\pi/2)}^{\pi/2} e^{s \sin \theta} d\theta. \quad (21)$$

Then

$$\chi^*(y) = \chi(s(y)) - ys(y), \quad \frac{d\chi(s)}{ds} = y, \quad \frac{d\chi^*(y)}{dy} = -s. \quad (22)$$

One may check from this that $\chi^*(y)$ and its first derivative are 0 at $y = 0$ and the second derivative is $-1/2$. It follows as above that

$$\gamma_5(n) = 4 \int_0^1 e^{n\chi^*(y)} dy \leq 4\sqrt{\pi} n^{-1/2} \approx 7.09 n^{-1/2}. \quad (23)$$

A lower bound was given at (9). In this case the crossing probability prediction falls below the bound (7).

What we have presented here is the stuff with which probability bounds and estimates can be made. The computer simulations provide an order of magnitude feel for crossings, but no more than this for n large. The reader desiring a more

precise recipe for “ n -noodle π ” may want to give more careful consideration to the small n cases, beginning with $n = 2$. We only hope to have whet the appetite for more probability and noodles.

REFERENCES

1. Adhikari, A. and J. Pitman (1989): The Shortest Plane Arc of Width 1, *American Math. Monthly*, 309–327.
2. Bhattacharya, R. and E. Waymire (1990): *Stochastic Processes with Applications*, John Wiley and Sons, NY.
3. Billingsley, P. (1986): *Probability and Measure*, John Wiley and Sons, NY.
4. Durrett, R. (1991): *Probability: Theory and Examples*, Wadsworth and Brooks/Cole, Pacific Grove, CA.
5. Kahane, J-P. (1985): *Some Random Series of Functions*, 2nd ed, Cambridge University Press, NY.
6. Ramaley, J. F. (1969): Buffon’s Noodle Problem, *American Math. Monthly*, no. 8, 916–918.
7. Snell, J. L. (1988): *Introduction to Probability*, Random House/Birkhauser, Boston.

Department of Mathematics
Oregon State University
Corvallis, OR 97331
waymire@math.orst.edu

PICTURE PUZZLE (from the collection of Paul Halmos)



What can these men talk about?
 (see page 570)

JOHN EMERT has been at Ball State University since receiving his Ph.D. in mathematics at the University of Tennessee in 1989. In addition to current research interests in geometric topology and computational geometry, he is active in the integration of computer algebra systems into the departmental curriculum. His alternative life as keyboardist and composer helped support his collegiate studies and continues as a diversion; he currently has twenty published compositions.

ROGER NELSON did his undergraduate work at Wheaton College (Illinois) and received his Ph.D. at Michigan State University in 1976 under the direction of Kwung W. Kwun. He taught at Hartwick College before joining the faculty at Ball State University in 1979. His research interests are in geometric topology. He has recently been involved in undergraduate curriculum development.

KAY I. MEEKS received the B.S. and M.A. degrees from Central Michigan University and the Ph.D. from the University of Tennessee-Knoxville in 1989. She taught at the University of Tennessee-Chatanooga before coming to Ball State University, where she is an Assistant Professor. Her interests in teacher preparation and Mathematics education at the collegiate level keep her an active member of the MAA and the NCTM. Dr. Meeks' current projects include additional work in geometry and the integration of mathematics with other subject areas.

ED WAYMIRE is a professor of mathematics at Oregon State University. He obtained his PhD in mathematics from the University of Arizona in 1976. His research interests are in probability theory and its applications. As an undergraduate he participated in one of NSF's undergraduate research initiatives to study electrification mechanisms of thunderstorms in the physics department at Southern Illinois University Edwardsville. Through fate and fortune he has maintained regular interdisciplinary involvement with scientists in the hydrologic and atmospheric sciences during his career as a mathematician. The 5-string banjo is another favorite distraction.

DAVID CALLAN received his Ph.D. from Notre Dame under O. T. O'Meara in 1977. Since then he has taught at several institutions. He enjoys problem solving and his mathematical interests include matrix theory and combinatorics. An avid cycle tourist, he has bicycled across two continents.

MOSS SWEEDLER received his B.S. from M.I.T. in 1963 and two years later his Ph.D. Beyond enjoying tasting new foods, designing furniture, kitchens and recipes, making challenging repairs with epoxy, creating cartoons and graphic art with Gwyneth Williams, dancing with the Fire House Dance Company, professing Mathematics at Cornell University and helping to establish a nature preserve, my son, Moss, is Director of the Army Center of Excellence for Symbolic Methods in Algorithmic Mathematics which is concerned with the mathematics and computer science at the interface of computer science and mathematics. Maybe he's overcompensating for being short.

"Abstractness, sometimes hurled as a reproach at mathematics, is its chief glory and its surest title to practical usefulness. It is also the source of such beauty as may spring from mathematics."

—E. T. Bell

**Answer to Picture Puzzle
(p. 559)**

What except algebra: they are Irving Kaplansky, Dan Zelinsky, and Gerhard Kalisch. The picture was taken in 1969 in Honolulu.

NOTES

Edited by: John Duncan

On a Curious Property of Counting Sequences

Victor Bronstein and Aviezri S. Fraenkel

This note is dedicated to the memory of Professor Joseph Gillis who passed away on Nov. 18, 1993, in his 82nd year.

1. INTRODUCTION. A *counting sequence* \mathcal{S} is a sequence of sequences $\{S_i\}_{i=0}^{\infty}$ of positive integers. The sequence S_{i+1} is obtained from S_i by counting the number m_k of times an integer k occurs in S_i and writing down in S_{i+1} the pairs m_k, k in increasing order of k , for all k for which $m_k > 0$.

Example. Beginning with $S_0 = (1)$, and with $S_0 = (6, 7)$, the first few elements of the two resulting sequences \mathcal{S} are depicted in Table 1.

TABLE 1. Initial elements of the counting sequences
for $S_0 = (1)$ and $S_0 = (6, 7)$

1	6 7
1 1	1 6 1 7
2 1	2 1 1 6 1 7
1 1 1 2	3 1 1 2 1 6 1 7
3 1 1 2	4 1 1 2 1 3 1 6 1 7
2 1 1 2 1 3	5 1 1 2 1 3 1 4 1 6 1 7
3 1 2 2 1 3	6 1 1 2 1 3 1 4 1 5 1 6 1 7
2 1 2 2 2 3	7 1 1 2 1 3 1 4 1 5 2 6 1 7
1 1 4 2 1 3	6 1 2 2 1 3 1 4 1 5 1 6 2 7
3 1 1 2 1 3 1 4	5 1 3 2 1 3 1 4 1 5 2 6 1 7
⋮	⋮

As usual, \mathbb{Z}^0 and \mathbb{Z}^+ denote the set of nonnegative integers and the set of positive integers respectively. The *length* $|S_i|$ of $S_i \in \mathcal{S}$ is the number of elements of S_i , (counting multiplicities). Thus $|S_1| = 2$ for S_1 in the left column of Table 1, and $|S_1| = 4$ for S_1 in the right column. Any sequence of finite length is called *finite*. What is the asymptotic behavior of the sequences S_i ? Does the length and hence the elements of $\{S_i\}$, grow without bound?

A counting sequence \mathcal{S} is called *ultimately periodic* if there exist positive integers i_0 and p such that $S_i = S_{i+p}$ for all $i \geq i_0$. The smallest such p is called the *period* of \mathcal{S} , and the smallest such i_0 is called the *preperiod* of \mathcal{S} . For example, $S_0 = (2, 2)$ is ultimately periodic with $i_0 = 0$ and $p = 1$, i.e., it is periodic with period 1 from the beginning. Our purpose is to prove the following surprising fact.

Theorem. Beginning with any finite initial sequence S_0 , all the sequences $S_i \in \mathcal{S}$ have bounded length, and \mathcal{S} is ultimately periodic.

For example, continuing Table 1 further, we get Table 2, from which we see that periodicity begins with $i_0 = 12$ ($p = 1$) for \mathcal{S} in the left column and with $i_0 = 10$ ($p = 3$) for \mathcal{S} in the right column.

TABLE 2. A continuation of Table 1

	\vdots	\vdots	
	4 1 1 2 2 3 1 4	5 1 2 2 2 3 1 4 2 5 1 6 1 7	$i_0 = 10$
	3 1 2 2 1 3 2 4	4 1 4 2 1 3 1 4 2 5 1 6 1 7	
$i_0 = 12$	2 1 3 2 2 3 1 4	5 1 2 2 1 3 3 4 1 5 1 6 1 7	
	2 1 3 2 2 3 1 4	5 1 2 2 2 3 1 4 2 5 1 6 1 7	

2. THE PROOF. We begin with some notation and definitions. Let $S_i \in \mathcal{S}$ ($i \geq 1$). Then S_i consists of pairs of positive integers $m_j(i), f_j(i)$. The elements $m_j(i)$ are called *multipliers* and the elements $f_j(i)$ —*factors*. Note that $f_1(i) < f_2(i) < f_3(i) < \dots$. It is easy to see that $|S_{i+1}| \leq 2|S_i|$ with equality if and only if all elements of S_i are distinct. Thus S_0 finite implies S_i finite for every $i \geq 1$. The sum of multipliers of S_i is denoted by $M(S_i)$, i.e., $M(S_i) = \sum_{j \geq 1} m_j(i)$. If $S_i, S_j \in \mathcal{S}$ with $j \geq i$, then S_j is a *successor* of S_i . The maximum element of S_i is denoted by $\max(S_i)$, i.e., $\max(S_i) = \max_j(m_j(i), f_j(i))$.

We now collect a few simple properties of \mathcal{S} .

- (i) S_i ($i \geq 1$), has *even* length. (S_i consists of *pairs*.)
- (ii) For $i \geq 1$, $M(S_i) = |S_{i-1}|$. (The sum of multipliers of S_i counts exactly all the elements of S_{i-1} .)
- (iii) The sequence of lengths is nondecreasing: $|S_1| \leq |S_2| \leq \dots$, and also $\max(S_1) \leq \max(S_2) \leq \dots$. (If any integer k appears in S_i , it must also appear in S_{i+1} ($i \geq 0$). But we may have $|S_0| > |S_1|$.)
- (iv) If S_{i+1} contains some integer not appearing in S_i , then it appears in S_{i+1} as a multiplier—not as a factor. (The set of factors of S_{i+1} equals the set of all distinct integers of S_i .)
- (v) For $i \geq 1$, $f_n(i)$ is equal to the maximum element of S_{i-1} , i.e., $f_n(i) = \max(S_{i-1})$.

(vi) Let

$$S_i = (m_1(i), f_1(i), m_2(i), f_2(i), \dots, m_n(i), f_n(i)) \quad (i \geq 1). \quad (1)$$

Then $f_n(i) \geq n$ (since $1 \leq f_1(i) < f_2(i) < f_3(i) < \dots$). If $f_n(i) = n$ (and hence $f_j(i) = j$ for all $j \leq n$), then S_i is called *complete*.

Lemma 1. Let S_i have the form (1) ($i \geq 1$). Then

$$\max_{1 \leq j \leq n} m_j(i) \leq |S_{i-1}| - \frac{1}{2}|S_i| + 1 \leq f_n(i) + 1.$$

Proof: In view of (ii), $m_j(i)$ attains its largest value if $m_h(i) = 1$ for all $h \neq j$. Thus by (ii),

$$\begin{aligned} m_j(i) &\leq |S_{i-1}| - (n-1) = |S_{i-1}| - \frac{1}{2}|S_i| + 1 \\ &\leq 2n - n + 1 = n + 1 \quad (\text{by (iii)}) \\ &\leq f_n(i) + 1 \quad (\text{by (vi)}). \end{aligned}$$

■

Corollary. Let S_i be given by (1). Then $\max(S_i) \leq \max(S_{i-1}) + 1$ ($i \geq 2$). If equality holds, then $\max(S_i) = \max_{1 \leq j \leq n} m_j(i) = f_n(i) + 1$.

Proof: If $\max_j m_j(i) < f_n(i) + 1$, then by (v), $\max(S_i) = f_n(i) = \max(S_{i-1})$. By Lemma 1, the only other possibility is $\max_j m_j(i) = f_n(i) + 1$. Then by Lemma 1 and by (v),

$$\max(S_i) = \max_j m_j(i) = f_n(i) + 1 = \max(S_{i-1}) + 1. \quad \blacksquare$$

Lemma 2. Let S_i be given by (1) with $n > 2$. Then $|S_j| \leq 2(f_n(i) + 1)$ for all $j > i$.

Proof: Suppose S_i has a successor S_j of length $2l > 2(f_n(i) + 1)$. Then (vi) implies that $f_i(j) > f_n(i) + 1$, hence the maximum element of S_j is also $> f_n(i) + 1$. By the Corollary, there exists a successor $S_k = (m_1(k), f_1(k), m_2(k), f_2(k), \dots)$ of S_i , with smallest index $k \geq i$, whose largest element is $f_n(i) + 1$, say $|S_k| = 2t$. The minimality of k and (iv) imply that $f_n(i) + 1$ is a multiplier of S_k , say $m_h(k) = f_n(i) + 1$, and $f_i(k) = f_n(i)$. Hence

$$M(S_k) = f_n(i) + 1 + \sum_{j \neq h} m_j(k) \geq f_n(i) + 1 + (t - 1) = f_n(i) + t.$$

On the other hand, by (ii) and (vi), $M(S_k) = |S_{k-1}| \leq 2t \leq t + f_n(i)$. The last two chains of inequalities imply that there is equality throughout. Thus

- a. $t = f_n(i) = f_i(k)$ and thus S_k is complete.
- b. For all $j \neq h$, $m_j(k) = 1$.

It follows that

$$S_{k+1} = (f_n(i), 1, 1, 2, 1, 3, \dots, 1, f_n(i) - 1, 1, f_n(i), 1, f_n(i) + 1),$$

$$S_{k+2} = (f_n(i) + 1, 1, 1, 2, 1, 3, \dots, 1, f_n(i) - 1, 2, f_n(i), 1, f_n(i) + 1),$$

with $|S_{k+1}| = |S_{k+2}| = 2(f_n(i) + 1)$.

Let S_q be a successor of S_{k+2} with smallest index q , which has an element $f_n(i) + 2$. Then $q > k + 2$. By (iv), $f_n(i) + 2$ appears in S_q as a multiplier, so S_{q-1} must have $f_n(i) + 1$ identical multipliers. Since $q - 1 \geq k + 2$, we have $|S_{q-1}| = 2(f_n(i) + 1)$. Hence all the multipliers of S_{q-1} have the same value, say r . By (ii), $M(S_{q-1}) = (f_n(i) + 1)r = |S_{q-2}|$. Now $q - 2 \geq k + 1$, so $|S_{q-2}| = 2(f_n(i) + 1)$, hence $r = 2$, so

$$S_{q-1} = (2, 1, 2, 2, 2, 3, \dots, 2, f_n(i), 2, f_n(i) + 1).$$

It follows that S_{q-2} has the numbers $1, 2, \dots, f_n(i) + 1$ as multipliers. Thus by (ii),

$$M(S_{q-2}) = (f_n(i) + 1)(f_n(i) + 2)/2 = |S_{q-3}| \leq 2(f_n(i) + 1),$$

so $f_n(i) \leq 2$. But $f_n(i) \geq n > 2$ by hypothesis. Thus there is no such successor S_q , and so $|S_i| \leq 2(f_n(i) + 1)$ for all $j > i$. \blacksquare

Proof of the Theorem: Lemma 2 implies that the sequences $S_i \in \mathcal{S}$ have bounded length. Since the number of sequences $S_i \in \mathcal{S}$ of length $|S_i| \leq 2k$ is also bounded, say by $(k + 1)^k$, the family \mathcal{S} is ultimately periodic. \blacksquare

If S_0 is an infinite sequence, it is possible that some successor is not defined. Thus if $S_0 = (1, 2, 3, \dots)$, then $S_1 = (1, 1, 1, 2, 1, 3, \dots)$ and $m(1)$ is not defined in

S_2 . But for $k = 1, 2, \dots$ we can let $m(k)$ be the smallest positive integer which makes S_0 a fixed point. Then

$$S_0 = (2, 1, 3, 2, 2, 3, 1, 4, 5, 5, 5, 6, 5, 7, 5, 8, 6, 9, 6, 10, 6, 11, \\ 6, 12, 7, 13, 7, 14, 7, 15, 7, 16, 8, 17, 8, 18, 8, 19, 8, 20, \\ 9, 21, 9, 22, 9, 23, 9, 24, 9, 25, 10, 26, \dots). \quad (2)$$

We could replace the prefix $p = (2, 1, 3, 2, 2, 3, 1, 4)$ of (2) by $p = (1)$, and then S_{12} would have the form (2).

QUESTIONS. (i) Give formulae or meaningful bounds for the preperiod i_0 . (As for the period p , it turns out that for the case of complete sequences, $p \leq 2$ except that $p = 3$ for the example in the right column of Table 1. A similar result holds for sequences that are not complete.)

(ii) Is there an infinite sequence S_0 for which every successor is defined, such that S_{i+1} differs from S_i in infinitely many elements for some i ?

(iii) Is there an infinite sequence S_0 such that every successor is defined and \mathcal{S} is not ultimately periodic?

ACKNOWLEDGMENT. We heard about counting sequences from J. Gillis who told us that they were proposed by John Conway.

*Department of Applied
Mathematics & Computer Science
The Weizmann Institute of Science
Rehovot 76100, Israel
victor@wisdom.weizmann.ac.il*

*Department of Mathematics
University of Pennsylvania
Philadelphia, PA 19104-6395
Permanent address:
Department of Applied
Mathematics & Computer Science
The Weizmann Institute of Science
Rehovot 76100, Israel
fraenkel@wisdom.weizmann.ac.il*

Chaos Without Nonperiodicity

Carsten Knudsen

Recently the mathematical definition of chaos proposed by Devaney in his book [3] has been revisited by several authors [1, 2]. The definition of chaos proposed by Devaney is

Devaney's definition of chaos. *Let f be a continuous transformation of a bounded metric space (X, d) . If f is topologically transitive, the periodic points of f are dense in X , and f exhibits sensitive dependence on the initial conditions, then f is said to be chaotic.*

Sensitive dependence on the initial conditions is from a physical point of view the central element of chaos, in fact, Gulick in his book [4] has defined chaos

synonymously with sensitive dependence on the initial conditions. Banks *et al.* [2] pointed out that sensitive dependence on the initial conditions is a redundant element in Devaney's definition, because it follows from topological transitivity and denseness of the periodic points. Assaf and Gadbois [1] later showed by construction of counter examples that neither topological transitivity nor denseness of the periodic points follow from the remaining two properties.

The aim of this note is to prove that sensitive dependence on the initial conditions and topological transitivity are stable properties under closure, as well as under restriction to dense invariant subsets. The consequence of the results to be presented is that chaos according to Devaney [3], may exist on bounded but non-compact spaces without any nonperiodic orbits.

Theorem 1. *Let f be a continuous transformation of a bounded metric space (X, d) . Let $f: Y \rightarrow Y$ where $\bar{Y} = X$. Then $f: X \rightarrow X$ is topologically transitive, if and only if $f: Y \rightarrow Y$ is topologically transitive.*

Proof of Theorem 1: Assume that f on X is topologically transitive. Let U and V be arbitrary open sets in X . $k \in \mathbb{N}$ exists such that $f^k(U) \cap V \neq \emptyset$. Let $U' = U \cap Y$ and $V' = V \cap Y$. Since U' is dense in U , then due to continuity of f , $f^k(U')$ is dense in $f^k(U)$, and finally invariance of Y ensures that $f^k(U') \cap V' \neq \emptyset$.

Assume that f restricted to Y is topologically transitive. Let U and V be arbitrary open sets in X , and let $U' = U \cap Y$ and $V' = V \cap Y$. $k \in \mathbb{N}$ exists such that $f^k(U') \cap V' \neq \emptyset$, and hence $f^k(U) \cap V \neq \emptyset$. \square

Theorem 2. *Let f be a continuous transformation of a bounded metric space (X, d) . Let $f: Y \rightarrow Y$ where $\bar{Y} = X$. Then $f: X \rightarrow X$ exhibits sensitive dependence on the initial conditions, if and only if $f: Y \rightarrow Y$ exhibits sensitive dependence on the initial conditions.*

Proof of Theorem 2: We first show that sensitive dependence on the initial conditions extends to the invariant subset Y . A sensitive dependence constant $\delta > 0$ exists such that in any neighborhood $N(x)$ of any $x \in Y$, $y \in X$ and $k \in \mathbb{N}$ exist such that $d(f^k(x), f^k(y)) > \delta$. Continuity of f guarantees the existence of a sequence $(y_n) \in Y$ with the limit point y . In particular the sequence (y_n) can be chosen such that all elements are contained in an open ball located in the interior of $N(x)$. Since f is continuous, the sequence $(f(y_n))$ has the limit point $f(y)$. Then we can choose a y' from (y_n) arbitrarily close to x such that $d(f^k(x), f^k(y')) > \delta/2$.

Next we show that sensitive dependence on the initial conditions extends to the closure of Y . For $f: Y \rightarrow Y$ we have a sensitive dependence constant $\delta > 0$. Pick any $x \in X \setminus Y$, and let $N(x)$ be an arbitrary neighborhood of x . A point $x_1 \in N(x) \cap Y$ exists. $N(x)$ is a neighborhood of x_1 , and since $f: Y \rightarrow Y$ exhibits sensitive dependence on the initial conditions $x_2 \in N(x) \cap Y$ and $k \in \mathbb{N}$ exist such that $d(f^k(x_1), f^k(x_2)) > \delta$. The orbit of x cannot follow the orbit of both x_1 and x_2 , since they separate, and hence $d(f^k(x), f^k(x')) > \delta/4$ where $x' \in \{x_1, x_2\}$. \square

Let us describe some consequences of the two theorems. Take a dynamical system that is chaotic according to Devaney's definition. Consider the restriction of the dynamics to the set of periodic points, which is clearly invariant. Theorems 1 and 2 imply that the restricted dynamical system is topologically transitive, and that the system exhibits sensitive dependence on the initial conditions. Together

with the trivially fulfilled condition of denseness of periodic points, this implies that the system is chaotic. Due to the lack of nonperiodicity this is not the kind of system most people would consider labeling chaotic. It may, of course, be argued that since for any periodic point there is another periodic point close by with arbitrarily high period, and therefore this is a kind of chaos. Nevertheless, insisting on the distinction between a truly nonperiodic orbit, say, a dense orbit, and a periodic orbit of even arbitrarily high period, such a system should not be termed chaotic. As another example let us consider a system from which we have removed all periodic points as well as their preimages, for instance the logistic map $x \mapsto 4x(1 - x)$ on the unit interval. According to Theorems 1 and 2 the restricted map is topologically transitive, and it exhibits sensitive dependence on the initial conditions. This system, on the other hand, is full of nonperiodic orbits and consequently deserves the labeling chaotic although the system has no periodic points.

As a consequence of the results presented in this note we therefore propose the following definition of chaos which excludes chaos without nonperiodicity, but allows, for instance, the example of the restricted logistic map to be considered as chaotic:

Proposed definition of chaos. *Let f be a continuous transformation of a bounded metric space (X, d) . If f has a dense orbit in X and f exhibits sensitive dependence on the initial conditions, then f is said to be chaotic.*

REFERENCES

1. D. Assaf, IV and S. Gadbois, Definition of chaos, *American Mathematical Monthly* 99 (1992) 865.
2. J. Banks, J. Brooks, G. Cairns, G. Davis, and P. Stacey, On Devaney's definition of chaos, *American Mathematical Monthly* 99 (1992) 332.
3. R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Addison-Wesley, 1985.
4. D. Gulick, *Encounters with Chaos*, McGraw-Hill, 1992.

*Physics Department and Center
for Chaos and Turbulence Studies
Bldg. 309
The Technical University of Denmark
DK-2800 Lyngby, Denmark
carsten@chaos.fl.dth.dk*

A Reverse Stolarsky's Inequality

Josip Pečarić

K. B. Stolarsky [1] proved the following result:

If $0 \leq g(x) \leq 1$ and g is nonincreasing on $[0, 1]$, then, for all positive numbers a and b , it follows that

$$\int_0^1 g(x^{1/(a+b)}) dx \geq \int_0^1 g(x^{1/a}) dx \int_0^1 g(x^{1/b}) dx. \quad (1)$$

Stolarsky also observed that by introducing the quotient $Q(g, p)$ as

$$Q(g, p) = \int_0^1 g(x) x^{p-1} dx \bigg/ \int_0^1 x^{p-1} dx,$$

we can make a change of variables and formulate (1) as

$$Q(g, a+b) \geq Q(g, a)Q(g, b). \quad (2)$$

A very simple proof of Stolarsky's inequality is given in [2].

Moreover, if we only suppose that g is a nonnegative nonincreasing function, using the substitution: $g(x) \rightarrow g(x)/g(0)$, we get from (2)

$$g(0)Q(g, a+b) \geq Q(g, a)Q(g, b). \quad (3)$$

In this paper we shall prove a reverse inequality.

Theorem. *If g is a nonnegative nondecreasing function on $[0, 1]$, then for all positive numbers a and b , the reverse inequality in (3) is valid.*

Proof: Note that

$$Q(g, p) = p \int_0^1 g(x) x^{p-1} dx = p \frac{x^p}{p} g(x) \bigg|_0^1 - \frac{1}{p} \int_0^1 x^p dg(x)$$

i.e.

$$Q(g, p) = g(1) - \int_0^1 x^p dg(x). \quad (4)$$

Also, it is obvious that we have

$$0 \leq \int_0^1 x^p dg(x) \bigg/ \int_0^1 dg(x) \leq 1.$$

The following result is valid:

$$(P+Q)(P+QAB) \leq (P+QA)(P+QB) \quad (6)$$

whenever $A \leq 1$, $B \leq 1$ and $PQ \leq 0$. In fact, the inequality (6) follows from the following equality

$$(1-A)(1-B)PQ = (P+Q)(P+QAB) - (P+QA)(P+QB).$$

Without loss in generality we can suppose that $g(0) - g(1) = 1$. In view of (5) we can apply (6) with $P = g(1)$, $Q = g(0) - g(1) (= 1)$, $A = \int_0^1 x^a dg(x)$ and $B = \int_0^1 x^b dg(x)$, to find that

$$\begin{aligned} & g(0) \left(g(1) - \int_0^1 x^a dg(x) \int_0^1 x^b dg(x) \right) \\ & \leq \left(g(1) - \int_0^1 x^a dg(x) \right) \left(g(1) - \int_0^1 x^b dg(x) \right) = Q(g, a)Q(g, b). \end{aligned} \quad (7)$$

A classical integral Čebyšev inequality (see, for example, [3, pp. 239–293]) gives:

$$\int_0^1 x^a dg(x) \int_0^1 x^b dg(x) \leq \int_0^1 x^{a+b} dg(x). \quad (8)$$

By combining (7) and (8) we get

$$Q(g, a)Q(g, b) \geq g(0) \left(g(1) - \int_0^1 x^{a+b} dg(x) \right) = g(0)Q(g, a+b),$$

and the proof is complete.

REFERENCES

1. K. B. Stolarsky, From Wythoff's Nim to Chebyshev's inequality, *Amer. Math. Monthly* 98 (1991), 889–900.
2. D. H. Luecking, Without Commensurability, *Amer. Math. Monthly* 99 (1992), 668.
3. D. S. Mitrinović, J. E. Pečarić and A. M. Fink, *Classical and New Inequalities in Analysis*, Kluwer Acad. Publ., 1993.

Faculty of Textile Technology
Pierottijeva 6
41000 Zagreb, Croatia
pecaric@mahazu.hazu.hr

A Note on Some Irrational Decimal Fractions

A. McD. Mercer

Let $1 \leq a_1 < a_2 < \dots$ be a strictly increasing sequence of positive integers and write $\text{Dec}\{a_k\}$ to mean the decimal fraction $0.(a_1)(a_2)\dots$. Recently the following result was proved in [1]:

Theorem A. *If*

$$\sum_{k=1}^{\infty} \frac{1}{a_k} = \infty$$

then $\text{Dec}\{a_k\}$ is irrational.

Taking a_k to be the k th prime this theorem provides a new proof that the decimal $0.2357111317\dots$ is irrational (see also [2] and [3]). On the other hand, it does not throw any light on the nature of, say, the number $\text{Dec}\{k^2\}$. Our purpose in this note is to prove the following generalization.

Theorem 1. *If there is an integer $r \geq 0$ such that the series*

$$\sum_{k=1}^{\infty} \frac{k^r}{a_k} = \infty,$$

then $\text{Dec}\{a_k\}$ is irrational.

This theorem assures us, for example, that all of the numbers $\text{Dec}\{k^m\}$ $m = 1, 2, \dots$ are irrational.

It should also be noted that the proof of Theorem 1, specialized to the case $r = 0$, differs from that of Theorem A and so provides a new proof of that result.

Proof of Theorem 1: Let us suppose that $x = 0.(a_1)(a_2)\dots$ is a rational number which, written in the usual decimal form, reads $x = 0.b_1b_2\dots$ ($0 \leq b_k \leq 9$). Since x is rational its decimal expansion will be periodic (at least after a first block of digits). Let the period be p . To fix ideas we shall assume that the periodic behaviour commences immediately after the decimal point. (If this were not the case we could simply move the decimal point sufficiently to the right and treat the fractional part of the new number.)

Let N_k be the number of digits in a_k . Clearly $N_k \leq N_{k+1}$ for each k and since the left-most digit of a_k is at least 1 then

$$a_k \geq 10^{N_k-1} \quad (= c_k \text{ say}) \quad (1)$$

Let the values taken by the sequence $\{N_k\}_1^\infty$ be $V_1 < V_2 < \dots$ and let the value V_k be taken exactly m_k times. We note that:

$$(i) \quad \text{If } N_k = N_{k+1} \text{ then } c_{k+1} = c_k \text{ and if } N_k < N_{k+1} \text{ then } c_{k+1} \geq 10c_k \quad (2)$$

(ii) At most p consecutive N_k 's can have the same value. To see this we observe that $N_\nu = N_{\nu+1} = \dots = N_{\nu+p}$ ($= N$ say) would imply $a_\nu = a_{\nu+p}$ because pN is a period of the decimal representation of x . Hence

$$m_k \leq p \quad (3)$$

(iii) If the jump $N_k < N_{k+1}$ occurs at the k -values $k = k_1, k_2, k_3, \dots$ then $k_j = m_1 + m_2 + \dots + m_j$. (4)

Let $r \geq 0$ be an integer. According to (1) and (4) the partial sums of the series $\sum_{k=1}^\infty (k^r/a_k)$ do not exceed

$$\frac{m_1(m_1)^r}{c_1} + \frac{m_2(m_1 + m_2)^r}{c_2} + \frac{m_3(m_1 + m_2 + m_3)^r}{c_3} + \dots$$

and this series converges since by (2) and (3) it is dominated by the series

$$\frac{p^{r+1}}{c_1} \sum_{k=1}^\infty \frac{k^r}{(10)^{k-1}}.$$

Hence $\sum_{k=1}^\infty (k^r/a_k)$ is convergent. This completes the proof of the theorem.

REFERENCES

1. Norbert Hegyvari, On Some Irrational Decimal Fractions. *Amer. Math. Monthly*, 100 (1993), 779–780.
2. Hardy–Wright, *An Introduction to the Theory of Numbers*, 5th edition, Oxford, Clarendon Press, 1979.
3. Polya–Szego, *Problems and Theorems in Analysis II*, Springer-Verlag, 1976, (exercise 257.)

Department of Mathematics and Statistics.
University of Guelph.
Ontario. N1G 1J4.
Canada.
amercer@msnet.mathstat.uoguelph.ca.

UNSOLVED PROBLEMS

Edited by: **Richard Guy and Richard Nowakowski**

In this department the MONTHLY presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial or related results. Typescripts should be sent to Richard Guy, Department of Mathematics & Statistics, The University of Calgary, Alberta, Canada T2N 1N4.

A Possible Permanent Formula

David Callan

The *permanent* of an n -square matrix $A = (a_{jk})$ is defined by $\text{per } A = \sum_{\sigma \in S_n} \prod_{j=1}^n a_{j\sigma(j)}$. The sum of the permanents of all $\binom{n}{r}^2$ r by r submatrices of A is denoted by $\sigma_r(A)$. Equivalently, $\sigma_r(A)$ is the sum of all products of r elements from A , no two in the same row or column. Thus if J is the n by n matrix of 1's, $\text{per}(xJ + A) = \sum_{r=0}^n r! \sigma_{n-r}(A) x^r$, providing a sort of generating function for the $\sigma_r(A)$ [we take $\sigma_0(A) = 1$].

Unlike its cousin, the determinant, few simple explicit formulas are known for the permanents of specific matrices. One of these few is Scott's formula: for $S = (1/(\omega^{2j-2} - \omega^{2k-1}))_{jk}$ where $\omega = e^{i\pi/n}$, it is known [1, 2, 3] that $\text{per } S = (-1)^{(n-1)/2} n(1 \cdot 3 \cdot 5 \cdots (n-2))^2 / 2^n$ if n is odd, $= 0$ if n is even. The proofs all use a classical result of Borchardt [4] (unfortunately of rather limited scope) that converts the problem to an evaluation of determinants, and this evaluation relies on the tractability of circulant matrices. (The n by n matrix (a_{jk}) is circulant if it is constant along each of its n extended diagonals, that is, if a_{jk} depends only on the value of $k - j$ modulo n .)

Now let D denote the diagonal matrix of n th roots of unity, $\text{diag}(1, \omega^2, \omega^4, \dots, \omega^{2n-2})$, and consider the (circulant) matrix $C = i(J - 2DS)$, for which $c_{jk} = \cot([2(k-j)+1]\pi/2n)$. The following remarkably simple formula for $\text{per}(xJ + C)$ has been verified by the author using *Mathematica*® for n up to 7.

Conjecture. For C as above,

$$\begin{aligned} \text{per}(xJ + C) = n! \left[x^n + \frac{n}{2} x^{n-2} + \frac{n(n-2)}{2 \cdot 4} x^{n-4} \right. \\ \left. + \frac{n(n-2)(n-4)}{2 \cdot 4 \cdot 6} x^{n-6} + \dots \right] \end{aligned} \quad (1)$$

(The last term is $(n(n-2) \cdots 3/2 \cdot 4 \cdots (n-1))x$ if n is odd, and is simply 1 if n is even.)

Replacing x by $-i$ in (1) and using the identity (Exercise: prove it!)

$$\begin{aligned} 1 + \sum_{j=1}^{\lfloor n/2 \rfloor} (-1)^j \frac{n(n-2) \cdots (n-2j+2)}{2 \cdot 4 \cdots 2j} \\ = (-1)^{(n-1)/2} \prod_{j=1}^{\lfloor n/2 \rfloor} \frac{n-2j}{2j}, \quad n \geq 2 \end{aligned}$$

yields Scott's formula, which tends to make the conjecture credible. (In this identity both sides are zero if n is even.)

We can at least establish the vanishing of $\sigma_r(C)$ that is implied by (1) for odd r . The first row of C looks like $(c_1, c_2, \dots, c_m, 0, -c_m, -c_{m-1}, \dots, -c_1)$ [n odd] or $(c_1, c_2, \dots, c_m, -c_m, -c_{m-1}, \dots, -c_1)$ [n even]. A cyclic permutation of C 's columns does not affect $\sigma_r(C)$ or C 's circulant-ness, so the following proposition covers both cases.

Proposition 1. Suppose A is an n -square circulant matrix whose first row is $(0, a_1, a_2, \dots, a_m, -a_m, -a_{m-1}, \dots, -a_1)$ [n odd] or $(a_1, a_2, \dots, a_m, -a_m, -a_{m-1}, \dots, -a_1)$ [n even].

Then $\sigma_r(A) = 0$ for odd r .

Proof: Consider the set L of n^2 positions (or *places*) in the matrix A . Suppose we can exhibit a bijection $\psi: L \rightarrow L$ with the following two properties:

(i) If places l_1, l_2 in L are not on the same line of A [$\{\text{lines}\} = \{\text{rows}\} \cup \{\text{columns}\}$], then neither are $\psi l_1, \psi l_2$.

(ii) If a is the entry in place l , then $-a$ is the entry in place ψl .

Then we are done because ψ induces a mapping on the set of products that go into $\sigma_r(A)$ [by (i)] that is actually a permutation [since ψ is] and *sign-changing* if r is odd [by (ii)]. So everything cancels out. If n is odd, the operation of transposition clearly fits the bill (and happens to be an involution). If n is even, the operation to use is: transpose first, then drop down one place (the first row being considered the successor of the last row). In symbols, $\psi(j, k) = (k+1, j)$ [arithmetic modulo n]. The easy verification that ψ has the requisite properties is left to the reader.

For Scott's matrix S itself, an analogous formula is fairly easily established: $\text{per}(xJ + S) = n!x^n + \text{per } S$. This is equivalent to: $\sigma_r(S) = 0$ for $0 < r < n$, which follows on taking $\theta = \omega^{-2}$ and $B = \frac{1}{2}(J + iC)$ in the following proposition.

Proposition 2. Let B be a circulant n by n matrix, θ a primitive n^{th} root of unity, and $D = \text{diag}(1, \theta, \theta^2, \dots, \theta^{n-1})$.

Then $\sigma_r(DB) = 0$ for $0 < r < n$.

Proof: Let $Q_{r,n}$ denote the set of increasing sequences of r integers, $\alpha(1) < \alpha(2) < \dots < \alpha(r)$ contained in $[1, n]$, and for $\alpha, \beta \in Q_{r,n}$, let $A[\alpha|\beta]$ denote the r by r submatrix of A whose rows (resp. columns) are indexed by α (resp. β). Thus $\sigma_r(A) = \sum_{\alpha, \beta \in Q_{r,n}} \text{per } A[\alpha|\beta]$. Finally, let T denote the shift operator on $Q_{r,n}$, that is, for $\alpha = (\alpha(j))_{j=1}^r \in Q_{r,n}$, to obtain $T\alpha$ add 1 to each term in α and, if necessary, reduce modulo n and cyclically permute to ensure ascending order. For example, for $\alpha = (1, 2, 4) \in Q_{3,5}$, $T\alpha = (2, 3, 5)$ while for $\alpha = (1, 2, 4) \in Q_{3,4}$, $T\alpha = (1, 2, 3)$.

Define two sequences in $Q_{r,n}$ to be equivalent if some number of shifts transforms one to the other. Then $Q_{r,n}$ is clearly partitioned into equivalence classes, each of size dividing n . Let $(\alpha_i)_{i \in I}$ be a set of representatives of these equivalence classes, and let n_i denote the size of class i . Note that rn_i is always a multiple of n (to see this, add the entries of the identical lists (modulo n) α_i and $T^{rn_i}\alpha_i = \{\alpha_i(j) + n_i\}_{j=1}^r$). Due to the circulant nature of B , it is easy to see that $\text{per } DB[T\alpha|T\beta] = \omega^r \text{per } DB[\alpha|\beta]$ and so, since T is a bijection, $\sum_{\beta \in Q_{r,n}} \text{per } DB[T\alpha|\beta] = \sum_{\beta \in Q_{r,n}} \text{per } DB[T\alpha|T\beta] = \omega^r \sum_{\beta \in Q_{r,n}} \text{per } DB[\alpha|\beta]$. Hence $\sigma_r(DB) = \sum_{i \in I} \sum_{j=1}^{n_i} \sum_{\beta \in Q_{r,n}} \text{per } DB[T^j\alpha_i|\beta] = \sum_{i \in I} (\sum_{j=1}^{n_i} \omega^{rj}) \sum_{\beta \in Q_{r,n}} \text{per } DB[\alpha_i|\beta] = 0$, the parenthesized factor vanishing for $0 < r < n$.

REFERENCES

1. D. Svrtan, Proof of Scott's Conjecture, *Proc. Amer. Math. Soc.* 87(2) (1983), 203–207, MR 84b:15008.
2. R. Kittappa, Proof of a Conjecture of 1881 on Permanents, *Linear and Multilinear Algebra* 10 (1981), 75–82, MR 80i:15005.
3. H. Minc, On a Conjecture of R. F. Scott (1881), *Lin. Alg. and Appl.* 28 (1979), 141–153, MR 82i:15008.
4. Henryk Minc, *Permanents*, Encyclopedia of Mathematics and its Applications, Addison Wesley, 1978, 6–8.

Department of Statistics
University of Wisconsin-Madison
1210 W. Dayton St
Madison, WI 53706-1693
callan@stat.wisc.edu

Added in Proof. The author announces a proof of this conjecture. Details will appear elsewhere.

Teachers

Dear Editor:

re Hungerford's *Future Elementary Teachers* (January 1994).

Why is it never suggested that mathematics in the elementary schools, like music, should be taught by teachers who *specialize* in that technical discipline? It is highly unlikely we will be able to bring *all* elementary teachers up to the required level of competence. But it *is* possible we could find a *few* of these teachers with a real talent for mathematics and then train them.

J. H. C. Creighton
33 Garces Drive
San Francisco, CA 94132

PROBLEMS AND SOLUTIONS

Edited by:

Richard T. Bumby, Fred Kochman and Douglas B. West

Proposed problems should be sent to the MONTHLY PROBLEMS address given on the inside front cover. Please include solutions, relevant references, etc. Three copies are requested.

Solutions of published problems should arrive before November 30, 1994 at the MONTHLY PROBLEMS address given on the inside front cover. Solutions should be typed with double spacing, including the problem number and the solver's name and mailing address. Two copies suffice. A self-addressed postcard or label should be included if an acknowledgment is desired.

*An asterisk (*) after the number of a problem, or part of a problem, indicates that no solution is currently available. Partial solutions will be useful in such cases. Otherwise, the published solution is likely to be based on a solution which is complete and correct. Of course, an elegant partial solution or a method leading to a more general result is always useful and welcome. In addition, references to other appearances of MONTHLY problems or to solutions of these problems in the literature are also solicited.*

PROBLEMS

10389. *Proposed by Raphael M. Robinson, University of California, Berkeley, CA.*

Find all solutions of the equation

$$a_1^{a_2 \cdots a_m} = b_1^{b_2 \cdots b_n}$$

where $m \geq 1$, $n \geq 1$, the a_i and b_i are integers with $2 \leq a_i \leq 4$ and $2 \leq b_i \leq 4$, and $a_1 \neq b_1$.

10390. *Proposed by Ognian Enchev, Boston University, Boston, MA.*

A standard deck of 52 playing cards is arranged at random in 4 rows and 13 columns. Show that with finitely many transpositions of cards of the same value (e.g., 7♣ and 7♥, K♦ and K♠, and so on) all cards can be arranged in such a way that each column contains one club, one diamond, one heart and one spade.

10391. *Proposed by Emre Alkan (student), Bosphorus University, İstanbul, Turkey, and the editors.*

If a_1, a_2, \dots, a_n are real numbers with $a_1 \geq a_2 \geq \dots \geq a_n$, and if ϕ is a convex function defined on the closed interval $[a_n, a_1]$, then

$$\sum_{k=1}^n \phi(a_k) a_{k+1} \geq \sum_{k=1}^n \phi(a_{k+1}) a_k$$

with the convention that $a_{n+1} = a_1$.

10392. Proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada.

Determine the extreme values of

$$\frac{1}{1+x+u} + \frac{1}{1+y+v} + \frac{1}{1+z+w}$$

where $xyz = a^3$, $uvw = b^3$, and $x, y, z, u, v, w > 0$.

10393. Proposed by Jean Anglesio, Garches, France.

Show that

$$\int_0^\infty \frac{e^{-ax}(1-e^{-x})^n}{x^r} dx = \frac{(-1)^r}{(r-1)!} \sum_{k=0}^n \binom{n}{k} (-1)^k (a+k)^{r-1} \log(a+k)$$

where $a \geq 0$ and $1 \leq r \leq n$ (except for $d = 0$, $r = 1$).

10394. Proposed by Ignacy I. Kotlarski, Oklahoma State University, Stillwater, OK.

Let N, Z_1, Z_2, \dots be a sequence of independent random variables, where N follows the geometric distribution with probabilities

$$\mathbf{P}(N = n) = p(1-p)^{n-1}$$

for $n = 1, 2, \dots$ with $0 < p < 1$, while the Z_j are identically distributed complex random variables $Z_j = X_j + iY_j$ where (X_j, Y_j) have density

$$f_{Z_j}(x, y) = \begin{cases} \frac{a}{2\pi} |z|^{a-2} & \text{for } |z| < 1 \\ 0 & \text{for } |z| \geq 1 \end{cases}$$

where $z = x + iy$ and $a > 0$. Find the distribution of

$$W = Z_1 \cdot Z_2 \cdots Z_N.$$

10395. Proposed by F. G. Boese, Max-Planck-Institut für extraterrestrische Physik, München, Germany.

Show that a bit more than 45.6% of all zeros of

$$\zeta_3(s) = 1 + 2^{-s} + 3^{-s}$$

lie in the open halfplane $\Re(s) < 0$. More precisely, show that these zeros lie in the vertical strip $-1 < \Re(s) < 0$.

NOTES

Notes: (10391) This includes the special case in which $a_n \geq 0$ and $\phi(x) = x^c$ for integer $c \geq 1$, which may be studied by algebraic means. **(10395)** Although the series $\sum_n n^{-s}$, only converges for $\Re(s) > 1$, there are formulas relating the Riemann Zeta function in the critical strip $0 < s < 1$ to the partial sums $\zeta_n(s) = 1 + 2^{-s} + \dots + n^{-s}$ (see E. C. Titchmarsh, *The Theory of the Riemann Zeta-function*, chapter IV). On the other hand, there is no reason to expect that the set of all zeros of the partial sums would be related to the set of zeros of $\zeta(s)$. This problem gives a strong quantitative difference in the first non-trivial case.

SOLUTIONS

Positive Solutions of a Matrix Equation

6666[1991, 655]. *Proposed by Marcel F. Neuts, University of Arizona, Tucson, AZ.*

Suppose A and B are n by n matrices with positive real entries, suppose A is non-singular, and let α denote the largest eigenvalue of BA .

(a) Prove that there exists an n by n matrix X with positive entries satisfying

$$X = A + XBX \quad (*)$$

if and only if $\alpha \leq 1/4$.

(b) If $\alpha \leq 1/4$, prove that

(i) there is a solution X_0 of $(*)$ with positive entries such that if X_1 is any other solution of $(*)$ with positive entries, then all entries of $X_1 - X_0$ are non-negative;

(ii) the maximum eigenvalue of BX_0 is $(1 - \sqrt{1 - 4\alpha})/2$;

(iii) if X_1 is any solution of $(*)$ with positive entries other than the minimal solution X_0 , then the maximum eigenvalue of BX_1 is $(1 + \sqrt{1 - 4\alpha})/2$.

Solution by R. Holzsager, The American University, Washington D.C. The following properties of matrices with positive entries are used:

1. There is a positive eigenvalue, called dominant, which exceeds the absolute value of any other eigenvalue.
2. There is a unique (up to scalar multiplication) eigenvector with positive entries; it corresponds to the dominant eigenvalue.
3. If r is the dominant eigenvalue of the positive matrix \mathbf{M} , then the sequence $\langle r^{-n} \mathbf{M}^n \rangle$ converges.
4. If \mathbf{v} is any positive vector, and $t\mathbf{v} \leq \mathbf{M}\mathbf{v}$ (coordinate-wise) then $t \leq r$; if $t = r$, then \mathbf{v} is an eigenvector.

These facts are all part of Perron-Frobenius theory and can be found in F. R. Gantmacher, *The Theory of Matrices*, Vol. 2, Chelsea, 1960 (chapter XIII, section 2, pp. 53–66), or similar books.

Proof of (a). From $X = A + XBX$, $BX = BA + (BX)^2$. If v is the positive eigenvector of the positive matrix BX , and r is the dominant eigenvalue, then $BAv = BXv - (BX)^2v = (r - r^2)v$, so v is also a positive eigenvector of BA . By uniqueness, $\alpha = r - r^2$. Since r is real (in fact, positive), $\alpha \leq 1/4$.

Conversely, if $\alpha \leq 1/4$, then the series

$$\sum \binom{2n}{n} \alpha^n / (n+1)$$

converges to $(1 - \sqrt{(1 - 4\alpha)})/2\alpha$. Since $\alpha^{-n}(BA)^n$ is uniformly bounded,

$$\sum \binom{2n}{n} (BA)^n / (n+1)$$

converges to some matrix U . Multiplying series gives $U = I + BAU^2 = I + UBAU$, so $AU = A + AUBAU$, and $X = AU$ is our desired solution.

Proof of (b). i) Let X_0 be the solution AU that we just found. If X_1 is another solution with non-negative entries, then starting with $X_1 \geq 0$ (entry-wise), and repeatedly substituting in $A + X_1BX_1$, we get an inductive proof that X_1 exceeds every partial sum of the series $X_0 = A + ABA + 2ABABA + 5ABABABA + \dots$. Passing to the limit, gives $X_1 \geq X_0$.

ii) $BX_0 = BAU = \sum \binom{2n}{n} (BA)^{n+1} / (n+1)$. If v is the positive eigenvector of BA , then it is also an eigenvector of BX_0 , with eigenvalue $\sum \binom{2n}{n} \alpha^{n+1} / (n+1) = (1 - \sqrt{(1 - 4\alpha)})/2$. By uniqueness again, this is the dominant eigenvalue of BX_0 .

iii) If X_1 is a non-negative solution, and w is the positive eigenvector of BX_1 , with eigenvalue r , then w is also an eigenvector of $BA = BX - (BX)^2$, with eigenvalue $r - r^2$. By uniqueness, this must equal α , so r must equal $(1 \pm \sqrt{(1 - 4\alpha)})/2$. If the minus sign holds, then $X_1 - X_0$ is a non-negative matrix by (i), and annihilates the positive vector v ; it is therefore 0. Thus, if X_1 is distinct from X_0 , the plus sign applies.

Editorial comment. Some solvers used a normalized form of the problem in which it was necessary to assume that A was non-singular. The selected solution can be seen not to require that hypothesis, although not all solvers taking this approach noted that fact. Only Reinhard Wolf gave a method to prove the *existence* of solutions with the larger eigenvalue. He notes that the set of matrices X with non-negative entries satisfying $BXv = \lambda v$ is a non-empty compact convex set in the usual identification of n by n matrices with \mathbb{R}^{n^2} . When v is the positive eigenvector of BA and $\lambda^2 - \lambda + \alpha = 0$, this set is taken into itself by $X \mapsto A + XBX$. The Brouwer fixed point theorem then shows that the equation has a solution. The equation shows that any solution with non-negative entries must have all entries positive.

Solved also by D. Ž. Đoković (Canada), I. Kastanas, O. P. Lossers (The Netherlands), R. Wolf (Austria), and the proposer.

A Sequence of Squares

10211 [1992, 361]. *Proposed by Herbert S. Wilf, University of Pennsylvania, Philadelphia, PA.*

Choose integers a, b, c, d , and let $K = 2(b^2 + a^2d - abc)$. Show that every member of the sequence defined by $y_0 = a^2, y_1 = b^2$ and

$$y_{n+1} = (c^2 - 2d)y_n - d^2y_{n-1} + Kd^n \quad (n \geq 1)$$

is the square of an integer.

Composite solution I by Stephen M. Gagola, Jr., Kent State University, Kent, OH, and O. P. Lossers, University of Technology, Eindhoven, The Netherlands. Define a sequence of integers by $x_0 = a, x_1 = b$, and $x_{n+1} = cx_n - dx_{n-1}$ for $n \geq 1$. Then set

$$M_n = \begin{pmatrix} x_n & x_{n-1} \\ x_{n+1} & x_n \end{pmatrix}.$$

Notice that

$$M_{n+1} = \begin{pmatrix} 0 & 1 \\ -d & c \end{pmatrix} M_n,$$

and hence $\det M_{n+1} = d \det M_n$. An easy induction then establishes that for all $n \geq 1$,

$$x_n^2 - x_{n+1}x_{n-1} = \det M_n = d^{n-1} \det M_1 = Kd^{n-1}/2.$$

From the relation $(x_{n+1} + dx_{n-1})^2 = c^2x_n^2$, we have

$$\begin{aligned} x_{n+1}^2 &= c^2x_n^2 - 2dx_{n+1}x_{n-1} - d^2x_{n-1}^2 \\ &= c^2x_n^2 - 2d(x_n^2 - Kd^{n-1}/2) - d^2x_{n-1}^2 \\ &= (c^2 - 2d)x_n^2 - d^2x_{n-1}^2 + Kd^n. \end{aligned}$$

Since $x_0^2 = a^2 = y_0$ and $x_1^2 = b^2 = y_1$, it follows that $x_n^2 = y_n$ for all n .

Composite solution II by Ilias Kastanas, California State University, Los Angeles, CA, and R. Glenn Powers, Western Kentucky University, Bowling Green, KY. For indeterminates a and b and fixed integers c and d , consider the endomorphism T of $\mathbb{Z}[a, b]$ generated by $T(a) = b$ and $T(b) = bc - ad$. Note that

$$T(K) = 2[(bc - ad)^2 + b^2d - b(bc - ad)c] = 2(b^2 + a^2d - abc)d = Kd.$$

Also $T(y_0) = b^2 = y_1$ and $T(y_1) = (bc - ad)^2 = y_2$. Assuming $T(y_j) = y_{j+1}$ for all $j \leq n$, we see that

$$\begin{aligned} T(y_{n+1}) &= (c^2 - 2d)T(y_n) - d^2T(y_{n-1}) + T(K)d^n \\ &= (c^2 - 2d)y_{n+1} - d^2y_n + Kd^{n+1} \\ &= y_{n+2}. \end{aligned}$$

It follows by induction that

$$y_n = T^n(y_0) = T^n(a^2) = (T^n(a))^2$$

Editorial comment. Most solvers discovered the sequence of square roots given in Solution I. David Zeitlin observed that this may also be established using

techniques explored in general in his paper, “Power identities for sequences defined by $W_{n+2} = dW_{n+1} - cW_n$ ”, *Fibonacci Quarterly*, 3 (1965), 241–256. Francisco Bellot and María Ascensión López pointed out that the particular case with $a = b = d = 1$ and $c = 4$ was a problem from the 1987 Bulgarian Mathematical Olympiad. A solution and generalization to that problem appeared in *Crux Mathematicorum*, 16 (1990), 292–294.

Solved by 49 other readers and the proposer.

Nearest Integer Zeta Functions

10212 [1992, 361]. *Proposed by Seung-Jin Bang, Seoul, Korea.*

Let $a(n)$ be the integer closest to $\sqrt[3]{n}$. Evaluate $\sum_{n=1}^{\infty} a(n)^{-4}$.

Composite solution by all solvers. The value is $\pi^2/2 + \pi^4/23040$. Indeed, for $s > 3$, one has

$$\sum_{n=1}^{\infty} a(n)^{-s} = 3\zeta(s-2) + 4^{-s}\zeta(s).$$

Here, $\zeta(s)$ denote the Riemann zeta function. The values of $\zeta(2n)$ for positive integers are easily computed rational multiples of π^{2n} .

Proof: The expression to be evaluated may be written as $\sum_{r=1}^{\infty} f(r)r^{-s}$ where $f(r)$ is the number of positive integers n for which $\sqrt[3]{n}$ is closest to r . Thus $f(r)$ is the number of positive integer solutions of

$$\left(r - \frac{1}{2}\right)^3 < n < \left(r + \frac{1}{2}\right)^3.$$

(The endpoints may be excluded since they cannot be integers.) This is an interval of length $3r^2 + 1/4$, so it will contain either $3r^2$ or $3r^2 + 1$ integers (all positive if $r > 0$). The number is $3r^2 + 1$ only if $(2r + 1)^3 \equiv 1 \pmod{8}$, and this is true if and only if $r \equiv 0 \pmod{4}$. Thus

$$\begin{aligned} \sum_{n=1}^{\infty} a(n)^{-s} &= \sum_{r=1}^{\infty} f(r)r^{-s} = \sum_{r=1}^{\infty} (3r^2)r^{-s} + \sum_{m=1}^{\infty} (4m)^{-s} \\ &= 3\zeta(s-2) + 4^{-s}\zeta(s). \end{aligned}$$

Editorial comment. Jonathan M. Borwein & Leo C. Hsu, Rick Mabry & Keith Neu, Josef Roppert, Douglas B. Tyler, and B. M. M. de Weger considered the more general sums $S_N(s) = \sum_{n=1}^{\infty} a_N(n)^{-s}$ where $a_N(n)$ is the integer closest to $\sqrt[N]{n}$. The method employed above for $N = 3$ also gives $S_2(s) = 2\zeta(s-1)$ and $S_4(s) = 4\zeta(s-3) + \zeta(s-1)$. For larger N , Hurwitz zeta functions, $\zeta(a, s) = \sum_{n=0}^{\infty} (n+a)^{-s}$ appear, and only Chu and the team of Borwein and Hsu persisted to give any values of $S_N(n)$ with $N > 4$. The latter solution contains a proof that $S_N(n)$ is a polynomial in π whose coefficients are algebraic numbers whenever $n - N$ is an

odd integer. For example, with

$$Q_6 = \frac{170912 + 49928\sqrt{2}}{15} \quad \text{and} \quad Q_7 = \frac{246013 + 353664\sqrt{2}}{45},$$

one gets

$$S_5(6) = \frac{5\pi^2}{6} + \frac{\pi^4}{36} + \left(\frac{1}{945} - Q_6 \sqrt{1 - \sqrt{\frac{1}{2}}} \right) \frac{\pi^6}{4^{12}}$$

$$S_6(7) = \pi^2 + \frac{\pi^4}{18} + \frac{\pi^6}{2520} + Q_7 \frac{\pi^7}{2^{27}}.$$

Solved by 65 readers (including those cited) and the proposer. One incorrect solution was received.

A Ring with No Nilpotents

10215 [1992, 362]. *Proposed by Michael Barr, McGill University, Montreal, Quebec, Canada.*

Let R be an associative ring (not necessarily commutative or possessing a unit element) with no non-zero nilpotent elements. Suppose that r and s are two elements of R such that $r^d = s^d$ and $r^e = s^e$, where d and e are relatively prime positive integers. Show that $r = s$.

Solution by Pat Stewart, Dalhousie University, Halifax, Nova Scotia, Canada. Choose integers a and b such that $ad + be = 1$. Without loss of generality, we may assume that $a > 0$ and $b < 0$. Since $r^d = s^d$, we have $r^{ad} = s^{ad}$ and hence $r^{1-be} = s^{1-be}$. Also, since $r^e = s^e$, we have $r^{-be} = s^{-be}$. Thus there is a positive integer u such that $r^{1+u} = s^{1+u}$ and $r^u = s^u$.

Substituting for s^u , we see that $r^{1+u} = r^u s$, and hence $r^u(r - s) = 0$. Elements x, y of a ring with no non-zero nilpotent elements satisfy $xy = 0$ if and only if $yx = 0$. Hence $r(r - s)r^{u-1} = 0$. Continuing in this fashion, one obtains by induction that $(r(r - s))^u = 0$, and so $r(r - s) = 0$. Similarly, $s(r - s) = 0$. Hence $(r - s)^2 = 0$, which implies $r - s = 0$, as desired.

Editorial comment. Several readers noted that the term *reduced* is used for rings without nilpotent elements. D. D. Anderson sketched an argument to show that, in a reduced ring, if $a_1 \dots a_n = 0$ and σ is any permutation of $[1, \dots, n]$, then $a_{\sigma(1)} \dots a_{\sigma(n)} = 0$. Frank Schmidt noted that the problem follows from the *structure theorem* for reduced rings: every reduced ring is a subdirect product of domains. W. K. Nicholson made the same observation, including a reference to source of the theorem (Andrunakevič and Rjabuhin, Soviet Math. Doklady 9 (1968), 565–567, MR 37 #6320) and the proof by A. Klein (Canad. Math. Bull. 23 (1980), 495–496).

Solved also by D. D. Anderson, G. Behrendt (Germany), B. W. Brock, D. Caccia, D. Callan, R. J. Chapman (U. K.), T. C. Craven, E. Dobrowolski and N. Buck (Canada), N. J. Fine, E. A. Herman, M. Hongan (Japan), M. Juvan (Slovenia), S. Kanetkar, K. S. Kedlaya (student), J. F. Kennison, J. J. Kuzmanovich, C. Lanski, S. C. Locke, O. P. Lossers (The Netherlands), R. F. McCoart, Jr., A. Müller (France), W. K. Nicholson (Canada), F. Schmidt, R. Stong, E. T. Wong, University of Wyoming Problem Circle, and the proposer.

A Nonlinear Integer-Valued Iteration

10247 [1992, 781]. *Proposed by Cristian Turcu, London, U.K.*

For a fixed real number A , define a sequence $\{X_n: n \geq 0\}$ by

$$X_0 = 0 \quad \text{and} \quad X_{n+1} = \frac{3X_n - \sqrt{5X_n^2 + 4A^2}}{2} \quad \text{for } n \geq 0.$$

(a) For which A is the sequence X_n convergent?

(b) For which A are all $X_n \in \mathbb{Z}$.

Solution by Lev Wertheim (student), Novosibirsk State University, Novosibirsk, Russia. (a) We have $X_1 = -|A|$, and clearly $X_{n+1} \leq \frac{3}{2}X_n \leq 0$. So $\{X_n\}$ converges if and only if $A = 0$, in which case $X_n \equiv 0$.

(b) From the definition of $\{X_n\}$ we can conclude

$$(2X_{n+1} - 3X_n)^2 = 5X_n^2 + 4A^2$$

or

$$X_{n+1}^2 - 3X_{n+1}X_n + X_n^2 = A^2.$$

Replacing n with $n-1$ in this last equation yields also $X_n^2 - 3X_nX_{n-1} + X_{n-1}^2 = A^2$.

Thus X_{n-1} and X_{n+1} are both roots of the quadratic equation

$$t^2 - 3X_nt + X_n^2 - A^2 = 0.$$

If $A \neq 0$, then the inequality in the proof of (a) is strict, so X_{n-1} and X_{n+1} are distinct. Hence they are all of the roots so by Vieta's theorem

$$X_{n-1} + X_{n+1} = 3X_n. \quad (1)$$

From this it is immediate that if $X_{n-1}, X_n \in \mathbb{Z}$ then $X_{n+1} \in \mathbb{Z}$. Thus $X_n \in \mathbb{Z}$ for all n iff $A \in \mathbb{Z}$.

Editorial comment. As observed by many solvers, we have

$$X_n = -F_{2n}|A|$$

where $\langle F_n \rangle = \langle 1, 1, 2, 3, 5, \dots \rangle$ is the usual Fibonacci sequence. This follows at once from the above solution since (1) is obviously satisfied if $X_n = -F_{2n}|A|$, and one checks that in fact

$$X_n = -F_{2n}|A|$$

for the first few terms.

Readers massaged the original recursion into a variety of different forms, from which Fibonacci behavior was the recognized and the integer property deduced.

More general sequences of this type are discussed in Murray S. Klamkin, "Perfect squares of the form $(m^2 - 1)a_n^2 + t$ ", *Math. Mag.* 42 (1969), 111-113.

^{*} Solved by 62 readers (including those cited) and the proposer. In addition, three readers solved only part (a).

Collaborating editors: David F. Appleyard, Paul T. Bateman, Duane M. Broline, Barry W. Brunson, Frank S. Cater, Gulbank D. Chakerian, Underwood Dudley, Gerald A. Edgar, Michael A. Filaseta, Ira M. Gessel, Richard A. Gibbs, Jerrold R. Griggs, Douglas A. Hensley, John R. Isbell, Mourad E. H. Ismail, Murray Klamkin, Daniel J. Kleitman, Frederick W. Luttman, Frank B. Miles, Richard Pfiefer, Stephen L. Portnoy, J. O. Shallit, John Henry Steelman, Kenneth B. Stolarsky, David E. Tepper, Douglas B. Tyler, Daniel Ullman, and William E. Watkins.

REVIEWS

Edited by **Darrell Haile**

Indiana-University, Bloomington, IN 47405

Ideals, Varieties, and Algorithms. By David Cox, John Little, Donal O'Shea.
Springer-Verlag, New York, 1992, vii + 513, \$39.95.

Reviewed by **Moss Sweedler**

Pure mathematics has resisted computers longer than most branches of science. The advent of computers is finally, fundamentally transforming the practice of mathematics. Both pure and applied mathematics and mathematicians are experiencing the upheaval. The changes include:

- *The line between pure and applied mathematics is in flux.
- *Computers spur new areas of research and revive languishing areas.
- *Computers increase productivity or make productivity possible by providing new means for:
 - generating examples and testing conjectures;
 - recording and disseminating information;
 - conducting joint research among far removed researchers.
- *Algorithm development motivates theory.
- *Computers enhance teaching by providing:
 - vastly improved graphical presentation;
 - ability to present non-trivial examples in class;
 - ability to routinely assign non-trivial homework problems;
 - tools for students to analyze, experiment and play with the subject matter.

Areas of mathematics such as applied logic, algebraic geometry, combinatorics, commutative algebra, dynamical systems, and many others, owe their existence, rebirth or major new research initiatives to the influence of computers. Researchers in these areas use computers as an extension of hand calculation to compute examples which test conjectures and provide data to increase understanding and motivate theorems. Beyond computation, computers exert a profound influence on mathematical research. Researchers delve into the algorithms underlying computation. They find motivation for pure mathematics from analyzing existing algorithms and developing new algorithms. Significant developments in the creation of algorithms require significant developments in theory. This is a fertile source of new mathematical theory.

Ideals, Varieties, and Algorithms by David Cox, John Little and Donal O'Shea is among the first of a new breed of algebraic geometry books. It is oriented toward the coming era emphasizing computation within mathematics. Algebraic geometry books from previous computational eras do not take into account the theory, computational techniques and viewpoints developed in the interim. *Ideals, Varieties, and Algorithms* is an introduction to the computational study of ideals, varieties, modern algorithms and, importantly, the ways in which algorithms

motivate theory. Computation aside, *Ideals, Varieties, and Algorithms* is a fine introduction to the study of algebraic geometry and commutative algebra.

The most significant algorithmic developments for commutative algebra and algebraic geometry since the last era emphasizing computation within mathematics, or perhaps of all time, are in the body of theory and techniques which have developed around the seminal constructive results of Bruno Buchberger. *Buchberger theory* is an appropriate name for the area, in the same way that *Galois theory* refers to a body of theory and techniques developed around Galois' seminal work. To the extent possible, Buchberger theory allows one to manipulate polynomials in several variables much as one manipulates polynomials in one variable. There are 7 keys to Buchberger theory. The first three generalize familiar ideas from polynomials in one variable:

<u>Buchberger theory</u>	<u>polynomials in one variable</u>
1. suitable notion of <i>leading term</i>	highest degree term of a polynomial
2. <i>reduction</i> over a set of polynomials	synthetic division with multiple divisors
3. <i>Groebner basis</i> of an ideal	principal generator of an ideal

Buchberger named Groebner bases after his thesis adviser Wolfgang Groebner. Fourth is the *S-pair*, a polynomial resulting from two initial polynomials. *S* is for syzygy. Fifth and sixth are Buchberger's *S-pair test* and the resulting *Buchberger algorithm* for constructing Groebner bases. Starting with a set *S* of polynomials, one pair wise tests if all the *S*-pairs reduce to zero over *S*. If so, *S* is a Groebner Basis for the ideal it generates. If not, form the union of *S* with the non-zero reduction results and repeat using this augmented *S*. The process eventually terminates with a Groebner basis for the ideal generated by the original *S*. Seventh are *multiplicative term orders* which play a fundamental role in both the basic theory and the applications.

Buchberger theory is as fundamental and more elementary than Galois theory and should be taught in undergraduate algebra courses and first year graduate algebra courses. *Ideals, Varieties, and Algorithms* is an excellent introduction to Buchberger theory and other computational methods. Not only are the techniques presented clearly, they are presented in the appropriate geometric context.

In 1985 I learned about Buchberger's work from Barry Trager at IBM's Watson Research Center. I was stunned. Surely he must be mistaken or else these powerful and simple techniques, useful for both computation and proofs¹ and available since the mid sixties, would routinely be included in algebra courses and algebra journals. Instead they were virtually unknown among academic algebraic geometers and commutative algebraists. I started writing a Buchberger book at the minimal level to present the algorithms; namely, just assuming familiarity with polynomials in several variables. Several chapters into the book I realized that it would be impossible to present many applications and that the entire treatment suffered from the lack of geometric context. *Ideals, Varieties, and Algorithms* is the book I wish I had written.

Buchberger theory, other computational methods, and the related theory are becoming known to commutative algebraists and algebraic geometers because of their importance for computation and because they provide the basis for many new

¹Originally Galois theory was primarily computational. Now a days, proofs often make use of the fact that a Galois group exists and no computation is involved. The same can be done with Buchberger theory.

interesting problems. The computer algebra community realized the significance of this work before the general academic pure mathematics community. The computer algebra community involves a full spectrum from engineers to theoreticians. The problem solvers using computer algebra, the implementors, the algorithm developers and the theory developers work side by side, motivating and benefiting from each other's developments and problems. At best, the result is a robust, stimulating, productive scientific environment. At worst, over zealous theoreticians—*Puritans*—and over zealous algorithm developers—*Algos*—divide into narrow-minded, self-serving communities. Each thinks it is smarter and works harder than the other. The *Algos* believe that algorithmic solutions to problems require better theorems with more difficult proofs than non-algorithmic solutions. The *Puritans* believe that the *Algos* work on problems which have already been solved.

Purely existential results are fundamental to algorithm development.² For implementation in computer algebra systems, one wishes to use efficient algorithms as measured by the number of computational steps and memory usage. Determining bounds on algorithms and finding efficient algorithms requires and motivates further theory. This is why algorithmic approaches to problems can be more difficult and involve more theory than non-algorithmic approaches.

Suppose a new area of mathematics has just opened. Here is a list of new problems and their difficulties.

Problem:

A	B	C	D	E	F ...
<i>Difficulty of non-algorithmic approach:</i>					
5	10	15	20	25	30 ...
<i>Difficulty of algorithmic approach:</i>					
6	12	18	24	30	36 ...

At first the *Puritans* and *Algos* work on problem A. An existential solution is obtained before an algorithmic solution. Soon the *Puritans* are working on Problem F, while the *Algos* are working on problem E. This is why the *Puritans* feel the *Algos* are working on problems which have already been solved. The *Algos* have contempt for problem F as being of *only theoretical* interest and irrelevant to the *real world*. The first *Algos* developing algorithmic approaches to F face this contempt from their *Algo* peers. Once F becomes established on an algorithmic basis, the *Algos* act as if they always believed it had *real world* relevance.

Given the antipathy between the two communities, it is not surprising that Buchberger's contribution is poorly understood and poorly acknowledged. The *Puritans* confuse the significance of the contributions of Buchberger and Hironaka. Buchberger's algorithm for constructing Groebner bases is the primary engine in modern commutative algebra and algebraic geometry. Groebner bases have an abstract characterization which is independent of any algorithm to create them.

² By the Hilbert basis theorem, an infinite generating set contains a finite generating set for ideals in polynomial rings over a field. Suppose we have a technique which produces a stream of polynomials (f_1, f_2, \dots) which generates an ideal I . I.e. for every $h \in I$ there are a finite number of polynomials $\{g_j\}$ where $\sum f_j g_j = h$. Furthermore, suppose we have a concrete test of whether a given finite set S generates I . Applying this test repeatedly to the sets $S_n = \{f_1, f_2, \dots, f_n\}$ we are assured by the Hilbert basis theorem that we reach a generating set for I . This insures we may terminate computation after a finite number of steps but gives no idea how many steps will be required.

This abstract characterization is the polynomial analog of *standard bases* for power series introduced by Hironaka, shortly before Buchberger's work. Consequently, Puritans ask:

Does Buchberger deserve credit for developing these bases for polynomial rings?

Should the bases in polynomial rings be called *Groebner* or *standard*?

Such questions miss the point and obscure the credit Buchberger deserves. Hironaka introduced no algorithm to create standard bases. The abstract notion of the basis without the S-pair test and resulting algorithm is like the glider rather than the airplane or the hot air without the balloon. While gliders rise to heights and have practical uses, compare the flight industry based on airplanes without engines with the flight industry based on airplanes with engines.

The first printing of the first edition of *Ideals, Varieties, and Algorithms* does not give Buchberger enough credit. It is easy to come away thinking: "Buchberger, he's the guy who invented standard bases after Hironaka did." It would have been refreshing and long overdue to come away thinking: "Buchberger, he's the guy whose fundamental work made a market for this book, provided an engine for a lot of computer algebra systems and is changing the face of commutative algebra and algebraic geometry." Cox, Little and O'Shea should say twenty Haile Bruno's and remedy this. Only twenty because they have fashioned a beautiful introduction to Buchberger theory.

For a more extensive catalog of algorithms in Buchberger theory, I recommend *Groebner Bases* by Becker and Weispfenning in the Springer GTM series. *Ideals, Varieties, and Algorithms* and *Groebner Bases* are excellent companions.

The authors and publisher are to be commended for holding down the price of *Ideals, Varieties, and Algorithms*. The availability of top quality, inexpensive books helps to advance the field. By no coincidence, Ruediger Gebauer at Springer worked in this area. The authors and publisher were magnanimous in letting the manuscript circulate quite freely up until publication. I imagine this policy helps the sale of good books.

Ideals, Varieties, and Algorithms appears in the Undergraduate Texts in Mathematics series. The mathematical prerequisites of the book as described in the preface are: "the students should have had a course in linear algebra and a course where they learned how to do proofs." I understand that the authors have successfully used the book as a textbook for a lower level undergraduate course. Nevertheless, I consider the book to be a wonderful text, not just a source book or reference book, for an upper level undergraduate or lower level graduate course. The exposition is very clear, there are many helpful pictures, and there are a great many instructive exercises, some quite challenging, providing a rich fund of homework problems. Applications made possible by computers are part of the impetus for the study and development of modern commutative algebra and algebraic geometry. The book presents applications in chapters on Robotics, Invariant Theory of Finite Groups and Automatic Geometric Theorem Proving. Automatic Geometric Theorem Proving is not an *artificial intelligence* attempt to duplicate the human theorem proving process. Instead, using analytic geometry, the theorem to be proved is reduced to a system of equations, possibly with some exceptional conditions or inequalities. One then uses algebraic means to verify the resulting algebraic system.

A number of books about computational commutative algebra and algebraic geometry are on the verge of publication. Many are strong on algebra and

algorithms and weak on the underlying geometry. The heart without the soul. *Ideals, Varieties, and Algorithms* offers the heart and soul of modern commutative algebra and algebraic geometry.

The review is based on the following. First contact was when the manuscript began circulating several years ago. Next contact was to review it for a publisher other than Springer. Recently, I conducted a reading course from the book. While I have tried to write impartially, it should be noted that I am friends with Bruno Buchberger, David Cox and Ruediger Gebauer.

Department of Mathematics
Cornell University
Ithaca, NY 14853-7901

Calculating for Cubics

Letter to the Editor:

Take a cubic equation $X^3 + pX + q = 0$ over a field F (of characteristic $\neq 2, 3$), and assume that its discriminant $D = -4p^3 - 27q^2$ is a square $D = d^2$ in F . The roots x_1, x_2, x_3 then are either in F or in a cyclic Galois extension of degree 3. Either trivially or by Galois theory, there is then a quadratic $Q(Y) = aY^2 + bY + c$ over F with $Q(x_1) = x_2$, $Q(x_2) = x_3$, $Q(x_3) = x_1$. Recently H. B. Griffiths and A. E. Hirst (MONTHLY 101, 151–161) derived a, b, c in terms of p, q as a consequence of other results, saying that “the algebraic theory does not explicitly tell us how to calculate [them].” Rising to the challenge, I shall sketch a straightforward computation.

Let $\zeta = (-1 + \sqrt{-3})/2$ be the cube root of 1. Set $s = -27q/2 + 3\sqrt{-3}d/2$, and let $u^3 = s$. Cardano's formulas give one root as $3x_1 = u - 3p/u = u - 3pu^2/s$; similarly, $3x_2 = \zeta u - 3p\zeta^2u^2/s$ and $3x_3 = \zeta^2u - 3p\zeta u^2/s$. Observe that replacing u by another cube root ζu or ζ^2u permutes the indices of the roots cyclically. Now the equation $Q(x_1) = ax_1^2 + bx_1 + c = x_2$ becomes $a(u^2/9 - 2p/3 + up^2/s) + b(-u^2p/s + u/3) + c = -p\zeta^2u^2/s + \zeta u/3$. If we just try to make the coefficients of $1, u, u^2$ agree, we get a system of equations for a, b, c :

$$\begin{aligned} a(-2p/3) + c &= 0, \\ a(p^2/s) + b/3 &= \zeta/3, \\ a(1/9) - b(p/s) &= -\zeta^2p/s. \end{aligned}$$

It is straightforward to solve this, and we get

$$\begin{aligned} a &= 9\sqrt{-3}ps/(s^2 + 27p^3) = 3p/d, \\ b &= \zeta - 9p^3/sd = -(9q + d)/2d, \\ c &= (2p/3)a = 2p^2/d. \end{aligned}$$

As u does not occur explicitly in these formulas, we can replace it by the other cube roots and find that these a, b, c also give $Q(x_2) = x_3$ and $Q(x_3) = x_1$.

William C. Waterhouse
Department of Mathematics
Penn State
University Park, PA 16802

TELEGRAPHIC REVIEWS

Edited by **Arnold Ostebee and Paul Zorn**

with the assistance of the Mathematics Departments of
Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T : Textbook	P : Professional Reading	1-4 : Semester
C : Computer Software	L : Undergraduate Library	** : Special Emphasis
S : Supplementary Reading	13 : Grade Level	?? : Questionable

Readers are advised that price information is subject to change. Selected books and software packages receive a second, more extensive review in the *Monthly*.

Books and software submitted for review should be sent to **Book Reviews Editor, American Mathematical Monthly, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098.**

General, T(13-14: 1), S, L*. *Concepts & Images: Visual Mathematics.* Arthur L. Loeb. Design Sci. Collection. Birkhäuser, 1993, xi + 228 pp, \$49.50. [ISBN 0-8176-3620-X] Non-technical introduction to mathematics of visual figures: symmetry, tessellation, glides, reflection, translation, quasi-symmetry, geometric construction, etc. In "workbook" style; many drawing and tracing exercises. Informal and inviting—a double antidote for visual illiteracy and math anxiety. PZ

General, S*(13-18), L.** *Spirals: From Theodorus to Chaos.* Philip J. Davis. AK Peters, 1993, ix + 237 pp, \$29.95. [ISBN 1-56881-010-5] Spirals in general and the Theodorus spiral in particular, from many, many points of view: historical, theoretical, computational, analytic, dynamic. Extensive notes and historical supplements. Generous, learned, entertaining; a *tour de force* in the liberal tradition. Based on author's 1990 Hedrick Lectures. PZ

Reference, S, C, P. *The Maple Handbook.* Darren Redfern. Springer-Verlag, 1993, 497 pp, \$29 (P). [ISBN 0-387-94054-5] Documents all commands available in Maple V Release 2. Concisely organized; 12 subject areas, each with a short introduction. DP

Mathematics Appreciation, S(15-16), P, L*. *The Broken Dice and Other Mathematical Tales of Chance.* Ivar Ekeland. Transl: Carol Volk. Univ of Chicago Pr, 1993, 183 pp, \$19.95. [ISBN 0-226-19991-6] A refreshingly literary look at the mathematics of luck and fate, told partly through Norse myths. BC

Recreational Mathematics, S(13). *Mathemat-*

ical Cavalcade. Brian Bolt. Cambridge Univ Pr, 1992, vi + 118 pp, \$17.95 (P). [ISBN 0-521-42617-0] 131 mathematical recreations (puzzles, curiosities, games, activities) on arithmetic, geometry, graph theory, logical thinking. No prerequisites (except high school algebra for some problems). LCL

History, P, L. *The Search for E.T. Bell, Also Known as John Taine.* Constance Reid. Spectrum Ser. MAA, 1993, x + 372 pp, \$35. [ISBN 0-88385-508-9] Engaging biography of one of the more colorful "men of mathematics" of the 20th century. BC

History, P, L*.** *A Century of Mathematics: Through the Eyes of the Monthly.* Ed: John H. Ewing. MAA, 1994, xi + 323 pp, \$39.50. [ISBN 0-88385-459-7] A wide-ranging and delightful collection of articles, excerpts, photos, notes, news clips, etc., from the *Monthly's* first 100 years (1894-1994). Mathematical exposition at its best, covering research, education, recreation, competitions, sociology, and more. A unique survey of modern American mathematics as discipline and as profession.

Logic, S(13-14), L.** *Introduction to Mathematical Philosophy.* Bertrand Russell. Dover, 1993, viii + 208 pp, \$6.95 (P). [ISBN 0-486-27724-0] An unaltered republication of the 1919 *Second Edition* of this classic. Topics include numbers, orders, relations, infinite cardinal numbers, limits and continuity, propositional functions, descriptions, and classes. DP

Foundations, T*(16-18: 1, 2), S*, P, L. *The Joy of Sets: Fundamentals of Contemporary Set Theory, Second Edition.* Keith Devlin. Under-

grad. Texts in Math. Springer-Verlag, 1993, x + 192 pp, \$29.95. [ISBN 0-387-94094-4] Following a brisk introduction to "naive" set theory, develops Zermelo-Fraenkel theory in detail: Z-F axioms, cardinal and ordinal numbers, axiomatic set theory, constructibility, etc. Final chapter outlines non-well-founded set theory, an alternative to Z-F. First half has exercises, problems; clear, reader-friendly exposition throughout. (Revised edition of *Fundamentals of Contemporary Set Theory*, TR, March 1980.) PZ

Group Theory, T*(17: 1). *Topics in Combinatorial Group Theory*. Gilbert Baumslag. Lect. in Math. Birkhäuser, 1993, vii + 164 pp, \$29.50 (P). [ISBN 0-8176-2921-1] Notes from a 1987-88 course at ETH Zürich. Topics include history, the Burnside problem, free groups and the Reidemeister-Schreier method, free products and HNN extensions, groups acting on trees. DP

Algebra, T(15-16: 1, 2). *Abstract Algebra: Theory and Applications*. Thomas W. Judson. PWS, 1994, xiii + 427 pp. [ISBN 0-534-93684-9] Groups, rings, fields, in that order. Sylow theorems, lattice and Boolean algebras, Galois theory, applications (e.g., cryptography, algebraic coding theory), and historical notes. LC

Complex Analysis, T(18: 1, 2), S, P. *Boundary Behaviour of Conformal Maps*. Ch. Pommerenke. Grund. der math. Wissenschaften, B. 299. Springer-Verlag, 1992, ix + 300 pp, \$69. [ISBN 0-387-54751-7] Wide-ranging survey of conformal mapping theory, starting with basics; stresses relation between geometry of simply-connected domains and analysis of their Riemann mapping functions. Each chapter begins with short, helpful overview. Readable exposition; many brief exercise sets; sizable bibliography. PZ

Differential Equations, T(14: 1, 2). *Elementary Differential Equations with Boundary Value Problems, Third Edition*. C.H. Edwards, Jr., David E. Penney. Prentice Hall, 1993, xv + 774 pp. [ISBN 0-13-253410-X] Flexible text: systems can be taught with or without assuming linear algebra; power series are reviewed before being applied. Impulses, Dirac delta functions are optional topics. Covers numerical techniques with Basic code. Plentiful exercises, many with solutions. (Second Edition, TR, December 1989.) SM

Differential Equations, P. *Global Behavior of Nonlinear Difference Equations of Higher Order with Applications*. V.L. Kocic, G. Ladas. Math. & Its Applic., V. 256. Kluwer Academic, 1993, xi + 228 pp, \$102. [ISBN 0-7923-2286-

X] Basics, global stability, rational recursive structures, applications, periodic cycles, open problems. Comprehensive bibliography. DH

Partial Differential Equations, P. *Asymptotic Behaviour of Solutions of Evolutionary Equations*. M.I. Vishik. Cambridge Univ Pr, 1992, 155 pp, \$39.95; \$18.95 (P). [ISBN 0-521-42023-7; 0-521-42237-X] Global approximate solutions of evolutionary equations are constructed from locally asymptotic solutions by perturbing the equations. Equations considered include Navier-Stokes, reaction-diffusion, and other systems of parabolic and hyperbolic type. 40-page appendix on non-autonomous systems. SP

Dynamical Systems, P. *Stochastic and Chaotic Oscillations*. Yu. I. Neimark, P.S. Landa. Math. & Its Applic., V. 77. Kluwer Academic, 1992, xii + 500 pp. [ISBN 0-7923-1530-8] Presents various dissipative dynamical systems (from engineering, mechanics, chemistry, and biology) that exhibit chaotic behavior. SP

Dynamical Systems, T(14: 1). *Mathematics of Models: Continuous and Discrete Dynamical Systems*. H. Brian Griffiths, Adrian Oldknow. Math. & Its Applic. Ellis Horwood, 1993, xiii + 435 pp, \$42. [ISBN 0-13-563800-3] Two parts: (1) introduction to classical models; geometric tools; more complex models (including catastrophe models, discrete nonlinear dynamics); (2) theoretical support. DH

Dynamical Systems, P. *Dynamical Systems VIII: Singularity Theory II, Applications*. Ed: V.I. Arnol'd. Ency. of Math. Sci., V. 39. Springer-Verlag, 1993, 235 pp, \$89. [ISBN 0-387-53376-1] Self-contained treatment of classification of functions and mappings, Legendre and Lagrangian singularities, Maxwell sets, gradient dynamical systems, singularities of the boundaries of domains of function spaces, ramified integrals, monodromy, deformation of real singularities, local lacunas, etc. SP

Dynamical Systems, P, L. *The General Problem of the Stability of Motion*. A.M. Lyapunov. Transl. & Ed: A.T. Fuller. Taylor & Francis, 1992, ix + 270 pp. [ISBN 0-7484-0062-1] English translation of Lyapunov's 1892 memoir. Develops notion of stability of equilibria, a central theme in dynamical systems. SP

Dynamical Systems, T(15-17: 2), L. *Dynamical Systems: Differential Equations, Maps and Chaotic Behaviour*. D.K. Arrowsmith, C.M. Place. Chapman & Hall, 1992, x + 330 pp. [ISBN 0-412-39070-1] For second DE course with qualitative emphasis. Treatment driven by examples. Many exercises—some require computer. Broad coverage. SP

Numerical Analysis, T(15: 1), S, C. *An Introduction to Numerical Linear Algebra.* Charles G. Cullen. PWS, 1994, xiii + 314 pp, disk included. [ISBN 0-534-93690-3] Numerical methods for solving linear systems directly and by iteration, finding eigenvalues and eigenvectors, linear least squares approximations. DH

Algebraic Topology, T(17-18: 2, 3). *Topology and Geometry.* Glen E. Bredon. Grad. Texts in Math., V. 139. Springer-Verlag, 1993, xiv + 557 pp, \$69. [ISBN 0-387-97926-3] Mainly devoted to algebraic topology with "glimpses into the beautiful and important realm of smooth manifolds." Covers the fundamental group, homology, cohomology, homotopy theory. DP

Operations Research, T(16-17), L. *Mathematics for Operations Research.* W.H. Marlow. Dover, 1993, xv + 483 pp, \$12.95 (P). [ISBN 0-486-67723-0] Reprint of 1978 Wiley edition (TR, December 1978). Topics include linear algebraic and differential systems, optimization in n dimensions. Terse exposition; assumes good calculus background. Many exercises and solutions, both theoretical and computational. SM

Probability, T(15-16: 1). *The Essentials of Probability.* Richard Durrett. Duxbury Pr, 1994, viii + 269 pp. [ISBN 0-534-19230-0] Concise, classical treatments of theory precede numerous examples and exercises. Final chapter briefly introduces statistical inference. RSK

Probability, T(14-15: 1). *Probability.* Jim Pitman. Texts in Stat. Springer-Verlag, 1993, xi + 559 pp, \$49. [ISBN 0-387-97974-3] Calculus-based but somewhat non-traditional introduction, stressing intuitive understanding and problem solving, not theory (e.g., no mention of moment generating functions on proof of the central limit theorem). Many advanced topics appear, particularly in exercises. RSK

Elementary Statistics, T*(13-15: 1, 2), L. *Statistics for Engineering Problem Solving.* Stephen B. Vardeman. PWS, 1994, xii + 811 pp. [ISBN 0-534-92871-4] Unusual introduction to subject. Stresses concepts needed by practicing engineers, from data collection and descriptive statistics to fractional factorial designs and mixture studies. Amply illustrated with real examples, MINITAB output. A final project integrates many of the methods covered. Extensive problem sets, many with real data. RSK

Elementary Statistics, T(13: 1). *Statistics in Practice.* Ernest A. Blaisdell. Saunders College, 1993, xxvi + 755 pp, \$52. [ISBN 0-03-032229-4] Well-written text on standard topics, with early introduction to regression analy-

sis, frequent exposure to P -values. MINITAB use is nicely interwoven. DH

Mathematical Statistics, T*(16-17: 2), L. *Statistical Theory, Fourth Edition.* Bernard W. Lindgren. Chapman & Hall, 1993, xii + 633 pp, \$57.95. [ISBN 0-412-04181-2] Revision and reorganization of 1976 *Third Edition* (TR, October 1976). Statistical decision theory now deferred until the end; Bayesian inference comes earlier. Forty percent more problems. Still a solid presentation of classical theory. RSK

Statistical Methods, T(16-17: 1), P. *Design of Experiments: A No-Name Approach.* Thomas J. Lorenzen, Virgil L. Anderson. Stat.: Textbooks & Mono., V. 139. Marcel Dekker, 1993, xi + 414 pp, \$59.75. [ISBN 0-8247-9077-4] Novel approach to experimental design. Avoids classical terminology in favor of four fundamental concepts: factorial designs, nested designs, restrictions on randomization (including randomized complete block and split plot designs), and fractional designs. Theoretical material and exercises are coded, can be skipped. Many coded, worked examples and key problems with solutions. RSK

Statistics, P*. *Bivariate Discrete Distributions.* Subrahmaniam Kocherlakota, Kathleen Kocherlakota. Stat.: Textbooks & Mono., V. 132. Marcel Dekker, 1992, xvi + 361 pp, \$125. [ISBN 0-8247-8702-1] Thorough introduction to the most prevalent bivariate discrete distributions, with latest techniques for computer simulation. Extensive bibliography, including references to multivariate discrete distributions. RSK

Programming, C, P. *C-XSC: A C++ Class Library for Extended Scientific Computing.* R. Klatte, et al. Transl: G.F. Corliss, et al. Springer-Verlag, 1993, xv + 269 pp, \$49.95 (P). [ISBN 0-387-56328-8] C-XSC (C-eXtended Scientific Computing) is a class library designed for very high numerical precision. Other features include data types for interval arithmetic and easy dynamic allocation of matrices. MPR

Computer Systems, P, L.** *UNIX for Programmers and Users: A Complete Guide.* Graham Glass. Prentice Hall, 1993, xxii + 633 pp, (P). [ISBN 0-13-480880-0] Excellent introduction and reference. From basic utility commands (ls, chmod, pwd, etc.), through complete discussions of the Bourne, Korn, and C shells. How to write scripts, how to use the Internet, and more. A must-have for regular UNIX users. With exercises, examples. MPR

Computer Graphics, T(16-17: 2). *L. Mathematics for Computer Graphics.* S.G. Hoggar. Tracts. in Theoretical Comp. Sci., V. 14. Cam-

bridge Univ Pr, 1992, xviii + 472 pp, \$39.95. [ISBN 0-521-37574-6] Mathematics underlying computer graphics: plane isometries, dihedral group, plane patterns, 3-space, matrices, isometries, other transformations, fractals, topology, iteration of complex functions. Scattered exercises; many illustrations. LC

Applications (Biological Science), P. *Population Dynamics and the Tribolium Model: Genetics and Demography*. Robert F. Costantino, Robert A. Desharnais. Mono. on Theoret. & Appl. Genetics, V. 13. Springer-Verlag, 1991, xii + 258 pp, \$89. [ISBN 0-387-97581-0] Interesting monograph integrates bifurcations, limit cycles, strange attractors, and stationary probability distributions with biological observations on flour beetles. Treats age structure dynamics, oscillations, natural selection, disequilibrium, cannibalism, coexistence. SP

Applications (Biological Science), P. L. *Theoretical Mechanics of Biological Neural Networks*. Ronald J. MacGregor. Neural Networks: Found. to Applic. Academic Pr, 1993, xii + 377 pp, \$69.95. [ISBN 0-12-464255-1] Aim is a foundation for neural network theory akin to Newton's foundation for classical mechanics. SM

Applications (Biological Science), P. *Fluid Dynamics in Biology*. Eds: A.Y. Cheer, C.P. van Dam. Contemp. Math., V. 141. AMS, 1993, xii + 586 pp, \$73 (P). [ISBN 0-8218-5148-9] 24 papers from a 1991 conference, including a very readable overview by James Lighthill, an article on computational biofluid dynamics by Peskin and McQueen, and one titled "Hairy little legs: Feeding, smelling, and swimming at low Reynolds numbers." BC

Applications (Communication Theory), P. L. *Wavelets: Algorithms & Applications*. Yves Meyer. Transl: Robert D. Ryan. SIAM, 1993, xi + 133 pp, \$19.50 (P). [ISBN 0-89871-309-9] Introduces wavelets, outlines history and a few applications: signal processing, computer vision, turbulence, study of galaxies, etc. JO

Applications (Engineering), P. *Artificial Intelligence in Engineering Design, Volume I: Design Representation and Models of Routine Design*. Eds: Christopher Tong, Duvvuru Sriram. Academic Pr, 1992, xv + 473 pp. [ISBN 0-12-660561-0] In two parts: knowledge representation, its effects on design and constraint satisfaction; frameworks for design problem solving, examples of designers. RJA

Applications (Engineering), L. *Software Systems for Structural Optimization*. Eds: H.R.E.M. Hörnlein, K. Schittkowski. ISNM, V. 110. Birkhäuser, 1993, viii + 283 pp, \$89.50. [ISBN 0-8176-2836-3] Finite element meth-

ods are used in structural design to test configurations for stresses, displacements, natural frequencies, etc. 11 software packages described here go one step further, producing feasible designs that optimize an objective function (e.g., weight or stiffness). SM

Applications (Engineering), S(17). *Methods of Engineering Mathematics*. Edward Haug, Kyung K. Choi. Prentice-Hall, 1993, xx + 585 pp. [ISBN 0-13-579061-1] Encyclopedic compilation of post-calculus mathematics needed for graduate work in engineering: linear algebra, analysis, mechanics, ODE's and PDE's, numerical methods. SK

Applications (Physics), T(17-18: 1, 2), S, P. *The Link Invariants of the Chern-Simons Field Theory: New Developments in Topological Quantum Field Theory*. Enore Guadagnini. Expos. in Math., V. 10. Walter de Gruyter, 1993, xiv + 312 pp, DM 148. [ISBN 3-11-014028-4] Self-contained text for introduction to topological quantum field theory. A detailed exposition of a non-trivial gauge theory that is exactly solvable in any 3-manifold. MU

Applications (Physics), S(15-16), L. *And Yet It Moves: Strange Systems and Subtle Questions in Physics*. Mark P. Silverman. Cambridge Univ Pr, 1993, xvii + 266 pp, \$49.95; \$24.95 (P). [ISBN 0-521-39173-3; 0-521-44631-7] Drawn from a rich professional career, salient examples of incongruities predicted by quantum theory are discussed, set in personal and historical context. Text is largely descriptive, but examples are buttressed by pertinent equations and charts. Excellent supplementary reading for a course in quantum mechanics. MU

Applications (Quality Control), P. *Reliability Improvement with Design of Experiments*. Lloyd W. Condra. Quality & Reliability, V. 41. Marcel Dekker, 1993, x + 370 pp, \$65. [ISBN 0-8247-8888-5] Practical guide for "reasonably intelligent and inquisitive technical personnel who wish to set up and operate product reliability programs," applying experimental design (primarily Taguchi methods) to improve reliability. Treats design of experiments, reliability methods, accelerated testing. With real examples from author's experience. RSK

Reviewers

RJA: Richard J. Allen, St. Olaf; LC: Laura Chihara, St. Olaf; BC: Barry Cipra, St. Olaf; DH: Deanna Haunsperger, St. Olaf; SK: Steve Kennedy, St. Olaf; RSK: Richard S. Kleber, St. Olaf; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; JO: Jeff Ondich, Carleton; SP: Samuel Patterson, Carleton; DP: David Peifer, St. Olaf; MPR: Matthew P. Richey, St. Olaf; MU: Milton Ulmer, Carleton; PZ: Paul Zorn, St. Olaf.

Bringing You Solutions.

Catalan's Conjecture

Are 8 and 9 the Only Consecutive Powers?

Paulo Ribenboim

Fully accessible and self-contained, this book provides a thorough historical study of the efforts of mathematicians to solve Catalan's problem. The book features a comprehensive bibliography and an appendix on Catalan's equation and powerful numbers, and includes many beautiful results of classical number theory not found in any other book.

February 1994, 364 pp., \$64.95
ISBN: 0-12-587170-8

Optimization Techniques in Statistics

Jagdish S. Rustagi

A Volume in the STATISTICAL MODELING AND DECISION SCIENCE Series

Statistics help guide us to optimal decisions under uncertainty. A large variety of statistical problems are essentially solutions to certain optimization problems. The mathematical techniques of optimization are fundamental to statistical theory and practice. In this book, Jagdish Rustagi provides full-spectrum coverage of these methods, ranging from classical optimization and Lagrange multipliers, to numerical techniques using gradients or direct search, to linear, nonlinear, and dynamic programming using the Kuhn-Tucker conditions or the Pontryagin maximal principle.

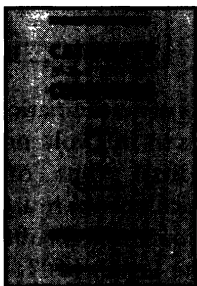
CONTENTS (Chapter Headings): Synopsis. Classical Optimization Techniques. Optimization and Inequalities. Numerical Methods of Optimization. Linear Programming Techniques. Nonlinear Programming Methods. Dynamic Programming Methods. Variational Methods. Stochastic Approximation Procedures. Optimization in Simulation. Optimization in Function Spaces. Chapter Exercises and Applications. Bibliography. Index.

April 1994, c. 352 pp., \$69.95
ISBN: 0-12-604555-0

The History of Mathematics

Volume III

Eberhard Knobloch and David E. Rowe



This volume contains nine essays that span three different approaches to the history of mathematics. It addresses the historiographical and philosophical issues involved in determining the meaning of mathematical history, and traces the convoluted development of the ideas of differential geometry and analysis.

The text also discusses the structure and interaction of mathematical communities through studies of these communities in the U.S. and in China.

May 1994, c. 450 pp., \$55.00 (tentative)
ISBN: 0-12-599663-2

Finite Elements for Analysis and Design

J. E. Akin

A Volume in the COMPUTATIONAL MATHEMATICS AND APPLICATIONS Series

This book provides a thoroughly revised and up-to-date account of the finite element method (FEM) and its numerous applications, with added emphasis on basic theory. Numerous worked examples are included to illustrate the material.

Paperback: \$39.95/ISBN: 0-12-047654-1
Casebound: \$79.00/ISBN: 0-12-047653-3
March 1994, 560 pp.
Includes a disk with FORTRAN code for the programs cited in the text.

From

Academic Press Journals

Historia Mathematica

Editor

David E. Rowe

Universität Mainz, Germany

Managing Editor

Karen Parshall

University of Virginia Charlottesville

Historia Mathematica is concerned with the history of all aspects of the mathematical sciences in all parts of the world and all historical periods. The journal publishes biographies of mathematicians and historians, studies of organizations and institutions, essays on historiography, and articles on the interactions among all facets of mathematical activity and other aspects of culture and society.

Published under the Auspices of the International Commission on the History of Mathematics of the Division of the History of Science of the International Union of the History and Philosophy of Science

Volume 21 (1994)

4 issues

ISSN 0315-0860

In the U.S.A. and

Canada: \$127.00

All other countries: \$156.00

Journal subscriptions are for the calendar year and are payable in advance.

Free sample copies are available on request. For more information, please contact:

Academic Press, Inc.

Journal Promotion Department

525 B Street, Suite 1900

San Diego, California 92101-4495, U.S.A.

1-800-894-3434

All prices are in U.S. dollars and are subject to change without notice.

Canadian customers: Please add 7% Goods and Services Tax to your order.



Order from your local bookseller or directly from

ACADEMIC PRESS, INC. A Division of Harcourt Brace & Company

Order Fulfillment Department DM17915, 6277 Sea Harbor Drive, Orlando, FL 32887

Call Toll Free 1-800-321-5068 or Fax 1-800-336-7377

Prices subject to change without notice ©1994 by Academic Press, Inc. All Rights Reserved. KRL/MW/MEH—02064

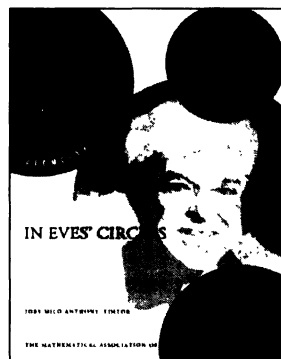
In Eves' Circles

Joby Milo Anthony, Editor

Howard Eves celebrated his eightieth birthday in 1991. To honor that occasion, the University of Central Florida sponsored a conference that focused on the lifelong interests of this prominent American mathematician namely, the history of mathematics, the teaching of mathematics, and geometry. Eves is well-known for his contributions to all three areas.

The conference was unique. Conference participants included pre-college mathematics teachers, community college and university teachers, and research mathematicians. Papers were delivered in sessions devoted to the classroom teacher, to the history of mathematics, and to pedagogical and research interests in geometry. Many lectures combined these subjects. This book presents some of those lectures. Anyone involved with teaching or producing mathematics can find something in this volume that will be interesting to them.

Some of these papers are specifically for the classroom teacher. They discuss a use of technology, or the organization of a class for some specific purpose. Other articles will provide teachers with examples of mathematical problems or historical episodes that can be used as part of a mathematics class. Still other papers deal with the



philosophy of mathematics education. There are articles in this volume that present new insights into some of the history of mathematics, and there are other articles that deal with some new results in geometry. This is really the legacy of Howard Eves. He has been both a mentor for teachers at every level, and a colleague of research mathematicians. And so some part of this volume will be appropriate for anyone interested in mathematics.

Also included in this volume is a penetrating interview with Eves.

220 pp., 1994, Paperbound

ISBN-0-88385-088-5

List: \$24.00

Catalog Number NTE-34

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Membership Code

Name _____

Address _____

City _____

State ____ Zip Code _____

Qty. Catalog Number Price

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

**Order
Today!**

OH NO!

THE LAST FREE ISSUE OF *MATH HORIZONS*



Minimum order is 20 copies with additional copies available in multiples of 10.

Copies	20	30	40	50	60	...	10n
Price	\$100	\$150	\$200	\$250	\$300	...	\$50n

**Don't disappoint your students.
Call 1-800-331-1622
to order your bulk
subscription!**

**Don't Miss
this
Opportunity!**

Research Issues in Undergraduate Mathematics Learning Preliminary Analyses and Reports

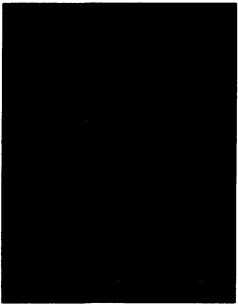
James J. Kaput and Ed Dubinsky, Editors

Research in undergraduate mathematics education is important for all college and university mathematicians. If our students are to be more successful in understanding mathematics, then college faculty need to understand how mathematics is learned. This knowledge can guide us in curriculum reform and in improving our own teaching. It can help us make mathematics accessible to all students and it can increase the number of graduate students in mathematics.

This volume of research in undergraduate mathematics education informs us about the nature of student learning in some of the most important topics in the undergraduate curriculum: sets, functions, calculus, statistics, abstract algebra and problem solving. Paying careful attention to the trouble students have in learning mathematics will help us to work with students so they can deal with those difficulties.

A survey of the literature begins the volume. Becker and Pence have brought together an unusually complete list of references on research in collegiate mathematics. Their comments will guide those attempting to begin or to continue a program of research in student learning.

The sad fact that even good calculus students stumble over nonroutine problems is the theme of Selden, Selden, and Mason. Their conclusions point to significant shortcomings in the curriculum. This study of student difficulties is



continued by Ferrini-Mundy and Graham who investigate a single student's interactions with the fundamental concepts of the calculus. Baxter studies a group of students to learn how they acquire the concept of set, while Cuoco does the same for the concept of function.

Cooperative learning does help the student. That is the conclusion of Bonsangue, who investigates how two carefully matched classes of students in a statistics course perform on exams. How students learn to write proofs in group theory is the subject considered by Hart. Rosamond breaks new ground by comparing how emotions vary in their effect on the problem solving ability of novices and experts.

All college faculty should read this book to find how they can help their students learn mathematics.

150 pp., Paperbound, 1994
ISBN 0-88385-090-7
List: \$24.00
Catalog Number NTE-33

ORDER FROM:
The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Membership Code _____

Name _____

Address _____

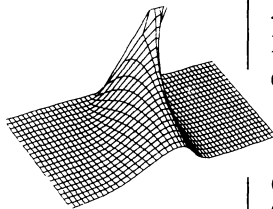
City _____

State _____ Zip Code _____

Qty.	Catalog Number	Price
Total \$		
Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
Credit Card No. _____		
Signature _____		
Exp. Date _____		

Applied Mathematics from SIAM

Society for Industrial and Applied Mathematics



Industrial Mathematics A Course in Solving Real-World Problems

Avner FRIEDMAN and Walter LITTMAN

"This is REAL. It has a different spirit. It gives students the distinct feeling that they could go into industry and actually work on problems like those in this book. The standard teaching of 'here is the mathematics, use it to solve this problem' has been replaced with 'here is a problem, use mathematics to solve it.' This book refreshes the interest of students in mathematics and motivates them to learn more of it. It helps them understand the nature and the importance of mathematics in real world applications."

— Oscar Bruno, Assistant Professor of Mathematics, Georgia Institute of Technology.

Contents

Introduction; Preface to the Student; Chapter 1: *Crystal Precipitation*; Chapter 2: *Air Quality Modeling*; Chapter 3: *Electron Beam Lithography*; Chapter 4: *Development of Color Film Negative*; Chapter 5: *How Does a Catalytic Converter Function?*; Chapter 6: *The Photocopy Machine*; Chapter 7: *The Photocopy Machine (Continued)*; Index.

Available July 1994 / Softcover / Approximately 152 pages / ISBN 0-89871-324-2

List Price \$22.50 / SIAM Member Price 18.00 / **Order Code OT42**

Special Classroom Adoption Price \$15.00 (minimum purchase of five copies)

Wavelets: Algorithms & Applications

Yves MEYER

translated and revised by Robert D. RYAN

"... I believe it is accessible to any scientifically minded reader with rudimentary knowledge of Fourier analysis; furthermore, the seasoned mathematician will find discussions of many interesting nonmathematical topics. ... I recommend the book as a delightful introduction to wavelets."

— Ingrid Daubechies, AT&T Bell Laboratories, *Science*, Vol. 262, December 3, 1993.

1993 / xi + 133 pages / Softcover / ISBN 0-89871-309-9

List Price \$19.50 / SIAM Member Price \$15.60 / **Order Code OT38**

TO ORDER

Use your credit card (AMEX, MasterCard and VISA): **Call toll free in USA: 800-447-SIAM**
Outside USA call: 215-382-9800 / Fax: 215-386-7999 / E-mail: service@siam.org
Or send check or money order to: SIAM, Dept. BKMA94, P.O. Box 7260, Philadelphia, PA 19101-7260

Payments may be made by wire transfer to SIAM's bank: PNC Bank, 3535 Market Street, Philadelphia, PA 19104; ABA Routing #031000053; Account Name: Society for Industrial and Applied Mathematics; Account Number 509-704-5.

Shipping and Handling: USA: Add \$2.75 for the first book and \$.50 for each additional book
Canada: Add \$4.50 for the first book and \$1.50 for each additional book. Outside USA/
Canada: Add \$4.50 per book. All overseas delivery is via airmail

Handbook of Writing for the Mathematical Sciences

Nicholas J. HIGHAM

Contents

Preface; Chapter 1: *General Principles*; Chapter 2: *Writer's Tools and Recommended Reading*; Chapter 3: *Mathematical Writing*; Chapter 4: *English Usage*; Chapter 5: *When English is a Foreign Language*; Chapter 6: *Writing a Paper*; Chapter 7: *Revising a Draft*; Chapter 8: *Publishing a Paper*; Chapter 9: *Writing a Talk*; Chapter 10: *Computer Aids for Writing and Research*; Appendix A: *The Greek Alphabet*; Appendix B: *Summary of TeX and LaTeX Symbols*; Appendix C: *GNU Emacs — The Sixty+ Most Useful Commands*; Appendix D: *Mathematical Organizations*; Appendix E: *Winners of Prizes for Expository Writing*; Glossary; Bibliography; Index.

1993 / xii + 241 pages / Softcover

ISBN 0-89871-314-5 / List Price \$21.50

SIAM Member Price \$17.20

SIAM Student Member Price \$12.00

Order Code OT39

Quantity discounts available for classroom use.

Mathematics Applied to Deterministic Problems in the Natural Sciences

C.C. LIN and L.A. SEGEL

Addresses the construction, analysis, and interpretation of mathematical models that shed light on significant problems in the physical sciences.

1988 / xxi + 609 pages / Softcover

ISBN 0-89871-229-7 / List Price \$28.00

SIAM Member Price \$22.40

Order Code CL01



For complete tables of contents or additional information about SIAM publications, access gopher.siam.org or contact SIAM Customer Service Department.

siam®

*Science and Industry
Advance with Mathematics*

Selected by CHOICE Magazine as one of the outstanding academic books of 1991.

More Mathematical Morsels

Ross Honsberger



Honsberger shows how powerful problem-solving is using elementary level mathematical techniques. His book contains a potpourri of intriguing and atypical problems and their solutions for the highly motivated student of mathematics...A wonderful source of interesting problems for highly gifted high school students, undergraduate mathematics students, and their instructors.

—CHOICE

Ross Honsberger is the best-selling author of seven books in the Dolciani Mathematical Exposition series, each of which presents problems from algebra, arithmetic, number theory, probability, and geometry, and provides ingenious solutions and/or intriguing results. His most recent addition to the series is **More Mathematical Morsels** which is a continuation of his best-selling **Mathematical Morsels** volume published in 1979.

All of the problems are accessible to anyone with a knowledge of freshman mathematics.

Here is Morsel 1: No matter which 55 positive integers may be selected from (1, 2,...,100), prove that you must choose some two that differ

by 9, some two that differ by 10, some two that differ by 12, and some two that differ by 13, but that you need not have any two that differ by 11.

315 pp., Paperbound, 1990

ISBN 0-88385-314-0

List: \$24.00 MAA Member: \$18.00

Catalog Number DOL-10

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Foreign Orders Please add \$3.00 per item ordered to cover postage and handling fees. The order will be sent via surface mail. If you want your order sent by air, we will be happy to send you a proforma invoice for your order.

Membership Code

Name _____

Address _____

City _____

State ____ Zip Code _____

Qty.	Catalog Number	Price
_____	_____	_____
_____	_____	_____
		Total \$ _____
Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
Credit Card No. _____		
Signature _____		
Exp. Date _____		

**No matter how you
express it, it still means
DERIVE® is half price.**

$$\begin{array}{ccccc}
 \lim_{x \rightarrow 0} \frac{1 - \cos x}{x^2} & \lim_{x \rightarrow 0} \frac{x}{\sin(2x)} & \frac{1}{2} \\
 50\% & & \\
 \sum_{n=1}^{\infty} \frac{1}{2^{n+1}} & 0.5 & \int_0^1 x \, dx
 \end{array}$$

DERIVE

The *DERIVE A Mathematical Assistant* program lets you express yourself symbolically, numerically and graphically, from algebra through calculus, with vectors and matrices too—all displayed with accepted math notation, or 2D and 3D plotting. *DERIVE* is also easy to use and easy to read, thanks to a friendly, menu-driven interface and split or

overlay windows that can display both algebra and plotting simultaneously. Better still, *DERIVE* has been praised for the accuracy and exactness of its solutions. But, best of all the suggested retail price is now only \$125. Which means *DERIVE* is now half price, no matter how you express it.

System requirements

DERIVE: MS-DOS 2.1 or later, 512K RAM, and one 3½" disk drive. Suggested retail price now **\$125 (Half off!)**.

DERIVE ROM card: Hewlett Packard 95LX & 100LX Palmtop, or other PC compatible ROM card computer. Suggested retail price now **\$125!**

DERIVE XM (eXtended Memory): 386 or 486 PC compatible with at least 2MB of *extended* memory. Suggested list price now \$250!

DERIVE is a registered trademark of Soft Warehouse, Inc.



Soft Warehouse
HONOLULU • HAWAII

Soft Warehouse, Inc. • 3660 Waiālae Ave.
Ste 304 • Honolulu, HI, USA 96816-3236
Ph (808) 734-5801 • Fax. (808) 735-1105

Winner of the MAA
Book Prize

The Mathematics of Games and Gambling

Edward Packel

The whole book is written with great urbanity and clarity...it is hard to see how it could have been better or more readable.

—The Mathematical Gazette

You can't lose with this **MAA Book Prize** winner if you want to see how mathematics can be used to analyze games of chance and skill. Roulette, craps, blackjack, backgammon, poker, bridge, state lotteries and horse races are considered here in a way that reveals their mathematical aspects. The tools used include probability, expectation, and game theory. No prerequisites are needed beyond high school algebra. This book is an excellent supplement to a course on probability.

No book can guarantee good luck, but this book will show you what determines the best bet in a game of chance or the optimal strategy in a strategic game. Besides being a good supplement in a course on probability and good bedside reading, this book's treat-

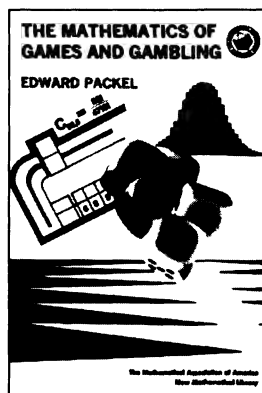
ment of lotteries should save the reader some money.

141 pp., Paperbound, 1981

ISBN 0-88385-628-X

List: \$16.00 MAA Member: \$13.00

Catalog Number NML-28



ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

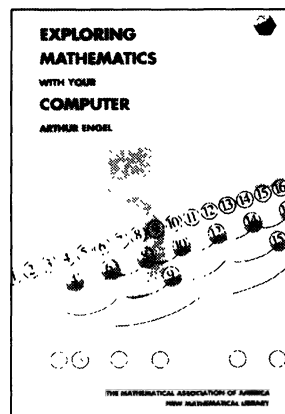
	Qty.	Catalog Number	Price
Membership Code -----			
Name _____			Total \$ _____
Address _____			Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
City _____			Credit Card No. _____
State _____ Zip Code _____			Signature _____
			Exp. Date _____

Exploring Mathematics With Your Computer

Arthur Engel

A must for academic libraries supporting an undergraduate major in mathematics. Public libraries with strong science collections should have this book as a resource for traditional mathematics topics and as a recreation for the mathematically inclined. Capable high school students would also benefit.

—CHOICE



Today's personal computer gives its owner tremendous power which can be used for experimental investigations and simulations of unprecedented scope, leading to mini-research. This book is a first step into this exciting field.

This is a mathematics book, not a programming book, although it explains Pascal to beginners. It is aimed at high school students and undergraduates with a strong interest in mathematics and teachers looking for fresh ideas. It is full of diverse mathematical ideas requiring little background. It includes a large number of challenging problems, many of which illustrate how numerical computation leads to conjectures which can then be proved by mathematical reasoning.

You will find 65 interesting and substantial mathematical topics in this book, and over 360 problems. Each topic is illustrated with examples and corresponding programs. The major goal of the book is to use the computer to collect data and formulate conjectures suggested by the data.

It is assumed that readers have a PC at their disposal.

264 pp., Paperbound, 1993

ISBN 0-88385-636-0

List: \$38.00 MAA Member: \$26.50

Catalog Number NML-35

A 3.5" IBM-compatible disk containing the Pascal programs described in the book is packaged with this volume.

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

Foreign Orders Please add \$3.00 per item ordered to cover postage and handling fees. The order will be sent via surface mail. If you want your order sent by air, we will be happy to send you a proforma invoice for your order.

Membership Code -----	Qty.	Catalog Number	Price
Name _____	_____		
Address _____	_____		
City _____	Total \$ _____		
State _____ Zip Code _____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
	Credit Card No. _____		
	Signature _____ Exp. Date _____		

Student Research Projects in Calculus

Marcus Cohen, Edward D. Gaughan, Arthur Knoebel,
Douglas S. Kurtz, and David Pengelley

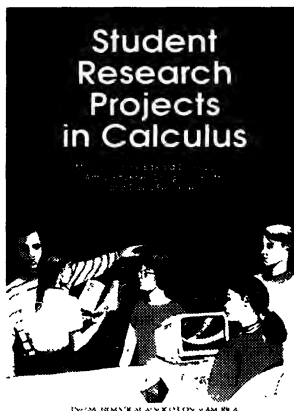
It is yet another workable way to rejuvenate college calculus. The monograph is highly recommended to those interested in experimenting with mathematics curricula.
—AAAS, Science Books & Films

I found the book very readable and thought provoking. Whether your students are in engineering, pure or applied science, or even liberal arts, this book may change the way you teach—and the way they learn—calculus for the better! —The Mathematics Teacher

An important contribution to the growing list of new curricular materials becoming available to assist in the teaching of calculus.
—CHOICE

Changing the way students learn calculus was the goal of the authors of this excellent guidebook. In the Spring of 1988, they began work on a student project approach to calculus.

You can use their methods in teaching your own calculus courses. Over 100 projects are presented, all of them ready to assign to your students in single and multivariable calculus. The projects were designed with one goal in mind: to get students to think for themselves. Each project is a multistep, take home problem, allowing students to work both individually and in groups.



Each project has accompanying notes to the instructor reporting students' experiences. The notes contain information on prerequisites, list the main topics the project explores, and suggest helpful hints. The authors have also provided several introductory chapters to help instructors use projects successfully in their classes and begin to create their own.

232 pp., Paperbound, 1992

ISBN 0-83385-503-8

List: \$25.50 MAA Member: \$18.00

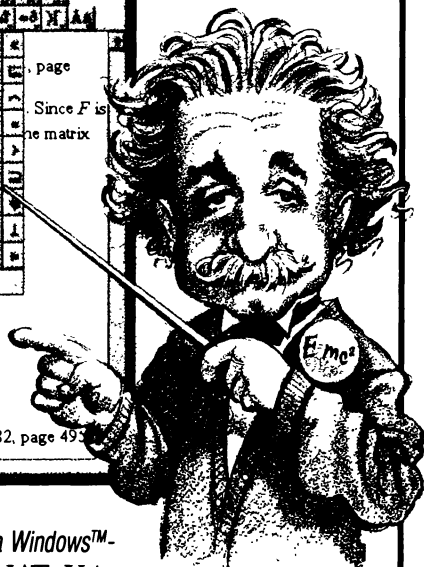
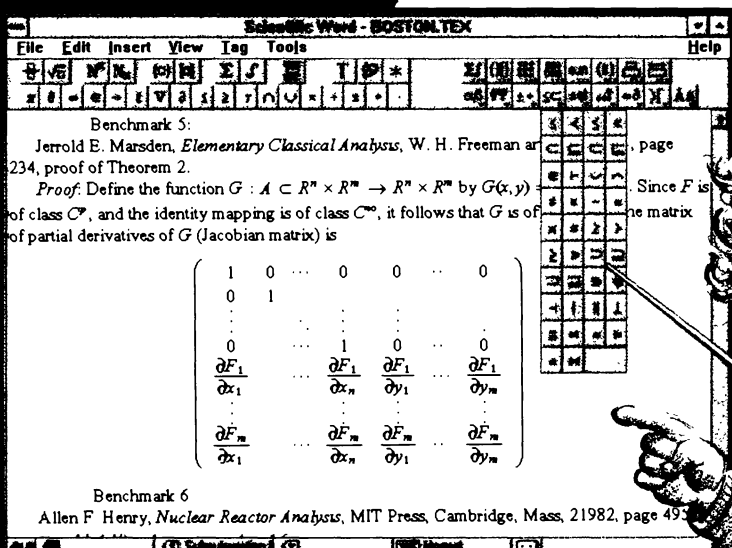
Catalog Number SRPC

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

		Qty.	Catalog Number	Price
Membership Code _____				
Name _____				
Address _____				
City _____				
State ____ Zip Code _____				
				Total \$ _____
		Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
		Credit Card No. _____		
		Signature _____		
		Exp. Date _____		

SCIENTIFIC[®] word



Discover the Genius of Scientific Word...

Scientific Word, a Windows[™]-based front-end to L^AT_EX is easy to learn and easy to use. It is a full document processor, not just an equation editor. You enter text and mathematics on a continuous screen without the distraction of popping in and out of equation boxes. With **Scientific Word** you use familiar mathematical notation to enter your mathematics – you need no special codes. **Scientific Word** adheres to internationally accepted mathematical formatting standards, so you are free to deal with the content of your document rather than its appearance. Your document is saved as an ASCII L^AT_EX file and printed output is produced via T_EX, the mathematical typesetting standard.

Call toll free **1-800-874-2383** to order your copy today. Ask about our 30-day money-back guarantee and our educational discount.

800-874-2383

CALL TODAY FOR MORE INFORMATION!

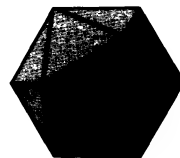
TCI
SOFTWARE RESEARCH

1190 FOSTER ROAD
LAS CRUCES, NM 88001
TEL: (505) 522-4600
FAX: (505) 522-0116
sales@tcisoft.com

SCIENTIFIC WORD IS A REGISTERED TRADEMARK OF TCI SOFTWARE RESEARCH.
T_EX IS A TRADEMARK OF THE AMERICAN MATHEMATICAL SOCIETY
WINDOWS IS A TRADEMARK OF MICROSOFT.

The American Mathematical Monthly

Volume 101 Number 6 / JUNE-JULY 1994
(ISSN 0002-9890)



Contents

ARTICLES

Juggling Drops and Descents / JOE BUHLER, DAVID EISENBUD,
RON GRAHAM, and COLIN WRIGHT 507

Teaching Integration by Substitution / DAVID GALE 520

Workable Gears, Archimedean Solids and Planar Bipartite Graphs /
GARY GORDON 527

On the Kummer Solutions of the Hypergeometric Equation /
REESE T. PROSSER 535

Reflections on a Mira / JOHN W. EMERT, KAY I. MEEKS,
and ROGER B. NELSON 544

Buffon Noodles / ED WAYMIRE 550

FEATURES

COMMENTS 506

PICTURE PUZZLE 559

NOTES

On a Curious Property of Counting Sequences / VICTOR BRONSTEIN
and AVIEZRI S. FRAENKEL 560

Chaos Without Nonperiodicity / CARSTEN KNUDSEN 563

A Reverse Stolarsky's Inequality / JOSIP PEČARIĆ 565

A Note on Some Irrational Decimal Fractions /

A. MCD. MERCER 567

THE AUTHORS 569

UNSOLVED PROBLEMS

A Possible Permanent Formula / DAVID CALLAN 571

PROBLEMS AND SOLUTIONS 574

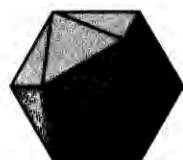
REVIEWS

Ideals, Varieties, and Algorithms. By David Cox, John Little,
and Donal O'Shea / MOSS SWEEDLER 582

TELEGRAPHIC REVIEWS 587



The American Mathematical Monthly



Volume 101, Number 7 / AUGUST-SEPTEMBER 1994



AN OFFICIAL PUBLICATION OF THE MATHEMATICAL ASSOCIATION OF AMERICA

NOTICE TO AUTHORS

The *Monthly* publishes articles, notes, and other features about mathematics and the profession. The readership of the *Monthly* is intended to include everybody who is mathematically inclined, including of course professional mathematicians and students of mathematics at all collegiate levels. While no single article or feature is likely to appeal to everyone, material should interest and be accessible to a large number of readers. This is the most important criterion for acceptance.

Articles may be expositions of old results or presentations of new ones. They may concern all of mathematics or one small area, a broad development or a single application, historical reminiscences or one important event. While some articles may contain the author's new research, the novelty of material and generality of the results is far less important than the clarity of exposition and general interest. Discussing one illuminating case of a well known result is far better than providing all the details of an obscure but new proposition. Articles in the *Monthly* are supposed to inform and to entertain; they are meant to be read rather than archived.

Notes are short and possibly informal articles. A note may concern a clever new proof of an old theorem, a novel way to present tired material, or a lively discussion of a philosophical (but still mathematical) issue. Also, any topic is suitable, so long as it is related to mathematics. Because a note is short, the first few sentences are the most important part: They should explain the purpose and invite the reader in. Photographs or diagrams often will attract the reader's attention.

All articles and notes should be sent to the editor:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Please send 3 copies, typewritten on only one side of the paper. Illustrations should be carefully drawn on separate sheets of paper in black ink; the original should be without lettering and two copies should have appropriate captions and lettering indicated.

Proposed problems or solutions should be sent to:

RICHARD BUMBY,
P.O. Box 10971
New Brunswick, NJ 08906-0971.

Please send 2 copies of all material, typewritten if possible.

Letters to the Editor, both for publication and for private reading, should be sent to the Editor at the address given above. Comments, including criticisms, are welcome, as are all suggestions for making the *Monthly* a lively, entertaining, and informative journal.

EDITOR:

JOHN H. EWING

ASSOCIATE EDITORS:

PETER BORWEIN	FRED KOCHMAN
RICHARD BUMBY	CATHERINE MCGEOCH
DENNIS DETURCK	RICHARD NOWAKOWSKI
UNDERWOOD DUDLEY	ARNOLD OSTEBEE
JOHN DUNCAN	LEE RUBEL
JOAN FERRINI-MUNDY	ABE SHENITZER
JOSEPH GALLIAN	LYNN STEEN
STEVEN GALOVICH	STAN WAGON
RICHARD GUY	DOUGLAS WEST
DARRELL HAILE	HERBERT WILF
PAUL HALMOS	SANDY ZABELL
JOAN HUTCHINSON	PAUL ZORN

EDITORIAL ASSISTANT:

MISTY CUMMINGS

STAFF ARTIST:

MIKE CAGLE

Reprint permission:

MARCIA P. SWARD, Executive Director

Advertising Correspondence:

Ms. ELAINE PEDREIRA, Advertising Manager

Subscription correspondence, change of address, and other inquiries:

Membership / Subscriptions Department

All at the address:

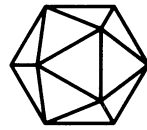
The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT. Copyrighted by the Mathematical Association of America (Incorporated), 1994, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source. Second class postage paid at Washington, DC, and additional mailing offices. **Postmaster:** Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.

**The American
Mathematical Monthly**

Volume 101 Number 7 / AUGUST–SEPTEMBER 1994
(ISSN 0002-9890)



Contents

ARTICLES

- A Tale of Two CD's / DAN KENNEDY 603
- Three Problems in Search of a Measure / JONATHAN L. KING 609
- The n -Queens Problem / IGOR RIVIN, ILAN VARDI, and
PAUL ZIMMERMANN 629
- What's the Difference Between Cantor Sets? / ROGER L. KRAFT 640
- Morphisms, Squarefree Strings, and the Tower of Hanoi Puzzle /
JEAN-PAUL ALLOUCHE, DAN ASTOORIAN, JIM RANDALL, and
JEFFREY SHALLIT 651

FEATURES

COMMENTS 602

NOTES

- Sierpinski's Theorem Is Deducible from Euler and Dirichlet /
A. A. AGEEV 659
- On Nonnegativity of Symmetric Polynomials / F. MATÚŠ 661
- New Tricks for Old Trees: Maps and the Pigeonhole Principle /
N. GRAHAM, R. C. ENTRINGER, AND L. A. SZÉKELY 664

THE COMPUTER SCIENCE SAMPLER

- Do You Know the Way to Vertex A ? / JEFF ONDICH 668

THE EVOLUTION OF...

- On the Calculus of Variations and Its Major Influences
on the Mathematics of the First Half of Our Century. Part I. /
ERWIN KREYSZIG 674

THE AUTHORS 679

PROBLEMS AND SOLUTIONS 681

REVIEWS

- How to Teach Mathematics*, By Steven G. Krantz /
MEYER JERISON 692

TELEGRAPHIC REVIEWS 695

COMMENTS

Teaching reform. My shelf is full of books and pamphlets produced by committees over the past 10 years. The titles are clever and catchy — about counting and assessing and measuring up. The presentations are slick (and expensive). The themes are uniformly gloomy, predicting disastrous consequences if we don't take action quickly to set things right. These pamphlets preach to politicians, not to mathematicians. They preach to the converted. Mostly, that's what committees do.

I am therefore delighted to find two wonderful books that directly or indirectly come from committees. The most recent is

You're the Professor WHAT NEXT

edited by Bettye Anne Case and published by the MAA. The other is

How to Teach Mathematics: A Personal Perspective

written by Steve Krantz and published by the AMS.

WHAT NEXT looks like a book written by a committee. It has a bit of everything — committee reports, survey results, case studies, solicited articles, and lists of resources. Nearly *half* the volume consists of appendices filled with reprinted short papers on career, social, and educational issues. Many of these may look familiar; they still make fascinating reading for mathematicians both young and old at every kind of institution. The theme of the book is simple: Young mathematicians ought to receive some organized professional training in graduate school. We ought to teach mathematicians how to be mathematicians. The book gives advice but no simple recipe for doing this (except to present ideas for professional seminars). Its vagueness is its strongest virtue. It is a book of ideas, not proclamations.

How to Teach was written as the result of a committee on which the author served. (A review by Meyer Jerrison appears in this issue of the MONTHLY.) It's organized into three sections, Guiding Principles (sample: Respect, Attitude, Math Anxiety, Computers), Practical Matters (sample: Voice, Eye Contact, Homework, Office Hours, Grading), and Sticky Wickets (sample: Non-Native English Speakers, Cheating, Discipline, Sexism). It's full of common sense advice about teaching in the real world — the kind of advice a senior colleague might communicate to a new instructor on the way to lunch or in the lounge over coffee. "Common sense", of course, does not mean "obvious" nor even "correct". But it's advice based on experience, not abstractions, and even when one disagrees with the author it's hard not to sympathize with his goals.

These books contain strong opinions and plenty of advice — but they do not preach disaster. Young mathematicians will profit from reading both from cover to cover. Older folks will profit even more because *both* books comment on everyday experiences we all share, and make us think harder about our profession.

Will you agree with everything in these books? Almost surely not. Will you be annoyed by some of the advice? Quite likely. Will they make you think about being a mathematician? That's guaranteed.

These books were written for mathematicians, not for politicians.

John Ewing

A Tale of Two CD'S

Dan Kennedy

These are indeed exciting times in the world of Mathematics. Stirred by the NCTM *Standards*, the winds of change are blowing through every level of the K-12 curriculum. At the same time, in colleges and universities the finest minds of our profession are turning their attention to the forging of a Calculus for a New Century. Earlier this year, we even saw the unexpected verification of Fermat's Last Theorem, something which surely none of us thought we would see in this lifetime. There is so much to talk about when mathematicians get together these days, yet, because we are teachers, we have so little time for talking. That is why I have chosen to write this article about none of these things.

Instead, I would like to write about records.

I have collected records since I was in grade school. By the time I was in graduate school, I owned more than two thousand 45's, three hundred albums, and miscellaneous 78's and EP's. I shared my hobby with friends, including some memorable years as music director and station manager of the college radio station at Holy Cross. I was one of those guys who could, upon hearing a golden oldie on the radio, quote the title, artist, and year of the song, and quite often the label and songwriter as well. My interest in the music of the moment naturally declined about the time disco music became popular, but by that time I had accumulated enough vinyl classics to keep myself and my party guests entertained forever. For example, one of my favorite ways to pass an evening with friends is to stage a "nostalgia playoff" between two guests, playing alternately the hit songs from their respective high school graduation years, until the quality of one year's hits is clearly unable to keep up with the quality of the other's. (In case you are interested, no guest has ever gone up against a graduate of 1957 without conceding after the 15th round or so.)

As a student of the recording industry, I would always sit up and take notice when some new product emerged which the prophets predicted would change the way people listened to music. The first big pretender to the vinyl throne was the 8-track tape. "The 8-track tape," they predicted, "will redefine the recording industry." It required no threading onto a spool, it did not scratch or shatter, it did not collect dust, it required no needle, it produced high fidelity sound for multiple speakers, and *you could play it in your car!* I was momentarily impressed, but I continued to buy records, and so did apparently a lot of other people. Today if you want to buy an 8-track tape you have to go to an antique show.

Then came the cassette tape. "The cassette tape," they predicted, "will refine the recording industry." It was smaller than the clumsy 8-track, but it had all its same advantages, including that of being playable in your automobile. Moreover, you could actually stick a few into your glove compartment. You could also buy a "portable tape player" which would play your cassettes on an arbitrary street corner at an arbitrary volume level. I myself continued to buy records, although I

did eventually buy a cassette recorder so that I could tape my records at home for playing later in my car. Although record stores eventually began selling albums on cassette, it was usually from a shelf toward the back of the store. Vinyl was still king.

Next came the laser disc. “The laser disc,” they predicted, “will redefine the recording industry.” The laser disc had the music encoded digitally, promising virtually perfect fidelity forever. It also would not scratch, smudge, or collect dust, and in place of the old diamond needle, which everyone always suspected would be fatal to plastic records eventually, there was a neat, powerful laser beam to lift the music off of the disc as cleanly as Scottie might beam up Captain Kirk. This was dazzling technology indeed; unfortunately, you could not play a laser disc in your car. For one thing, each disc was the size of a medium pizza; for another thing, you had to *sell* your car in order to buy a laser disc *player*, which cost several thousand dollars. Needless to say Hugh Hefner bought one for every room of his mansion, while the rest of us just kept buying records.

Then along came the compact disc. “The compact disc,” they predicted, “will redefine the recording industry.” And in an incredibly short period of time, it did.

Walk into a record store today and it probably won’t even be *called* a record store. The bins that once stored the vinyl now store row upon row of CD’s. Oh, the biggest stores will still carry a few records, but you often have to walk past the cassette section in order to find them. If you want a *real* adventure, try replacing your old phonograph needle! The recording industry has been completely taken over by the compact disc.

Why did the CD succeed where the other technologies had failed? Simply put, it was such a perfect idea that nobody who loved music could resist it. The sound was virtually perfect; the discs themselves were rugged; the players were affordable; and these things were, as their name implied, *compact*—you could fit dozens of them into a shoebox and carry them to the home of your friend. There, you would almost assuredly find another CD player. Everyone could share in the miracle.

Of course, for a while I held out. After all, I had this enormous investment in records, not to mention the means for playing them. But I would go the homes of my friends and hear the crystalline strains of CD music, and I would be jealous of that incredible *sound*. Finally, I realized that it was not the *records* that I liked; it was the *music*, and the music could be heard better on CD’s. I bought myself a player and began collecting compact discs. Most of my first compact disc purchases were actually albums that I already owned on vinyl, but I bought them so that I could rediscover them on a new level. Now I hardly ever play my records unless I am hosting a graduation year playoff. I still own them, but they are doing something that records do unfortunately well: They are gathering dust.

I have lived long enough to see the very essence of my lifelong hobby redefined.

But records are only my hobby; my profession is teaching mathematics. I suppose I became interested in mathematics at about the same time I started collecting records. I was in seventh grade when the Russians launched Sputnik, thereby kicking off some interesting times for my mathematics teachers. My high school courses were taught out of paperback textbooks authored by the School Mathematics Study Group, code SMSG, whose approach, they predicted, would redefine the way we taught and learned mathematics.

That was the New Math, and it lasted long enough to develop a reputation bad enough to spawn the Back to Basics movement. “The Back to Basics Movement,” they predicted, “will redefine the way we teach and learn mathematics.”

Then suddenly we all got distracted by computers. Computers were doing well in the stock market, and were pretty much running most companies and the government, so there was a strong feeling that we should all find out how they work. What most people discovered was that they worked by *mathematics*, which was enough for most people, because it meant that mathematics teachers could henceforth be held responsible for explaining the remaining details to their children. Computers were installed in many schools, because, they predicted, “computers will redefine the way we teach and learn mathematics.”

I was part of a three-man committee in 1978 that persuaded our Board of Trustees to sink 100 thousand dollars into a computer system that featured a Data General Nova 830 with 32K RAM, seven CRT terminals, and a teletype printer. Five years later we were back before them, hats in hand, pleading for another 100 thousand dollars to upgrade to a Hewlett-Packard 8000/30, plus ten more terminals and a sensible printer. Five years later we all but abandoned the HP and built a computer lab with shiny new Apple II's, at a cost of another 100 thousand dollars. Now we have an entire multimedia lab, fully stocked with Macintoshes, laser printers, scanners, CD-ROMs, and networking hardware, while our gleeful supplier has another 100 thousand of our school dollars. The amazing thing about this buying frenzy is that *each time* we pleaded our case with the Trustees, we assured them that *computers would redefine the way we taught and learned mathematics*. Incredibly, they fell for it every time.

But the sad truth of the matter was that we were still teaching and learning mathematics the way we had been doing it for decades. For all its marvelous capabilities, the computer was not changing the way that we taught and learned mathematics, and it was costing our school approximately 100 thousand dollars every five years to prove to ourselves that this was so.

Meanwhile, pocket scientific calculators had quietly appeared on the scene and had been welcomed in most mathematics classes. They *did* change a few things, but at such a mechanical level that people hardly noticed. Trig tables and log tables died a hasty and largely unmourned death, and now everyone had an equal chance at finding the purchase price of 4 CD's at \$16.95 a piece, after a 10% discount and a 7.5% sales tax have been figured in. To most teachers, this was the sort of inconsequential drudgery that machines were *supposed* to do. Indeed, such was the inflexibility of the mathematics curriculum that these machines were welcomed precisely *because* they made so little difference in what we taught and learned. Significantly, the only place where they were at all controversial was at the elementary level, where people were nervous about children losing the ability to multiply and divide on paper. In any event, pocket calculators did not redefine the way we taught and learned mathematics.

By now you have probably figured out my little parable. SMSG and Back to Basics may have shifted things a little bit, but only in terms of decades, while the inflexibility of the mathematics curriculum must be measured on a geological scale. New points of emphasis come and go periodically, denting the curriculum monolith with the same approximate impact as that of the 8-track tape on the recording industry.

Scientific calculators were nice, and they even got into the classrooms, but they were the cassettes of our profession. Sure, everyone has one, but cassettes never replaced the records, which were still the way that serious people played their music, and scientific calculators never replaced factoring, which was what students *really* needed to know if they wanted to succeed in serious mathematics.

Computers took us to the laser disc stage, and failed to redefine the way we taught and learned mathematics for the same reasons that laser discs failed to redefine the recording industry. It had nothing to do with the capabilities of the technology, and everything to do with the mood of the market place. No machine can inspire a revolution if Hugh Hefner is the only one who owns one.

Of course, there is another chapter in this story. We are still writing that chapter as I write these words, but I firmly believe that you and I have lived long enough to see the compact disc of our profession: an instrument which is so perfectly suited to what we do, that it is in the process of redefining the way we teach and learn mathematics. We call this wonderful machine a graphing calculator, but it is in fact a computer, with ironically the same computing power in kilobytes as that first computer system we bought at my school fifteen years and 400 thousand dollars ago. With this machine my students can do far more than compute; they can conjecture, they can model, and they can make connections—the very things that I want to teach them to do. Moreover, they *own* this technology; they do not have to go over to a rich friend's home or to a special room at school to use it. Nor do I even have to tell them how or when to use it. Like the compact disc, it has become a part of their lives.

Oh sure, for a while I tried to treat this technology the way I treated the cassettes in the record store. I bought one and used it, but I never thought it would redefine my profession. After all, I had twenty years of my life invested in the math curriculum monolith, and I had become pretty successful at teaching the traditional courses in some embarrassingly traditional ways. But I was open to change, and I had read the *Standards*, so I began to chip away at my preconceptions of what and how I had to teach. The first thing I did was to let them use their graphing calculators all the time. The next thing I did was to start every class with a problem, which the students would talk out until a solution emerged that they could explain to each other. What I discovered, of course, was how useless my crisp set of lecture notes had been all these years. The students were discovering the results *without me*, and then *showing each other how to solve the problems*. That left me free to walk around and answer questions. Every so often I still tie things together or generalize, but for the most part I let the course evolve through what the students are doing, and I provide the direction by the problems I select. I'm still not sure what the heck I'm doing, but I do know this: There is more mathematics going on in my classroom these days than there ever has been before. Now there are more people doing it.

Was the graphing calculator responsible for transforming my entire approach to teaching? Well, yes and no. What the graphing calculator did was get me to question what I had been doing for twenty years. It also got me focused on how I would get the students using it, which in turn got me focused on student learning rather than my own teaching. Eventually, after sacrificing the first few sacred cows, I acquired a taste for sacred beef and the rest was easy. And believe me, that same thing is going on in mathematics classrooms all across the country, in a movement that is growing exponentially.

When I was young, all of my radical friends were in reform school. Today, all of my radical friends are in school reform. It is a crazy, free-wheeling time, not unlike the political scene in eastern Europe, and it's really pretty exciting—once you overcome the initial sensation of being totally lost in a Brave New World. Let's face it, though: For years we mathematicians were in a rather unique position in the world of academia, being smugly certain that we could *all* teach the *exact* same curriculum in the *exact* same way because we knew *exactly* what was best for

our students. The other subject area committees in the AP program fought among themselves all the time, while my colleagues and I on the Calculus committee nodded sagely in unison, reviewing yet another problem about a region being rotated about the x -axis. It was a happy, homogeneous, unrealistic world, totally incompatible with the spirit of creative discovery that had characterized the evolution of mathematics since the dawn of cognition.

Did you ever wonder what Newton would say if he could come back today and watch a traditional calculus class in action—if you could call it “action”? Do you suppose he would be flattered to see that, 300 years after his death, we were all teaching *his same results* to our students? I don’t want to put words into the old Lion’s mouth, but knowing that Newton once wrote to Robert Hooke, “If I have seen further it is by standing on the shoulder of giants,” I dare say that he might pick up a graphing calculator, stare at it for several minutes with amazement, then say something like this: “The creators of this magic are giants. Is there nobody here who would stand on their shoulders?”

It is almost axiomatic in the academic world that the most creative minds belong to the mathematicians. Sadly, in the world of mathematics education that renowned creativity has been stifled for too long. To counter the trend, I tried a few years ago to release my own creative Muse in an evening of unbridled mathematical activity. I sat down with a fresh pad of paper and a pitcher of martinis, and before the evening was over and the martinis were gone I had produced 75 proofs of never-before-seen theorems. (Actually, I underachieved. It was an 80-proof gin.) I also created the perfect calculus problem, and here it is:

THE ALL-PURPOSE CALCULUS PROBLEM

©1993 Dan Kennedy

A particle starts at rest and moves with velocity $v(t) = \int_1^t e^{-x^2} dx$ along a 10-foot ladder, which leans against a trough with a triangular cross-section two feet wide and one foot high.

Sand is flowing out of the trough at a constant rate of two cubic feet per hour, forming a conical pile in the middle of a sandbox which has been formed by cutting a square of side x from each corner of an 8” by 15” piece of cardboard and folding up the sides.

An observer watches the particle from a lighthouse one mile off shore, peering through a window shaped like a rectangle surmounted by a semicircle.

- How fast is the tip of the shadow moving?
- Find the volume of the solid generated when the trough is rotated about the y -axis.
- Justify your answer.
- Using the information found in parts (a), (b), and (c), sketch the curve on the pair of coordinate axes given.

Okay, maybe I went too light on the vermouth. But here’s the sad part: Every teacher reading this article knew those problems by heart, because we’ve been teaching those same basic problems for years. That, indeed, is what calculus teachers have been doing for decades and decades. We have been teaching the math literate of tomorrow with the problems of yesterday, while explaining to them all the while that they will *need this mathematics in the future*. My friends, this is the stuff of which the Emperor’s New Clothes are made! While we have been spinning golden oldies on the phonograph, perhaps more accurately the victrola, the world of Mathematical Reality has gone CD.

Today, thankfully, all of that is changing. To cite just one close-to-home example of that, the College Board has given final approval to the AP Committee’s recommendation that the AP examinations be made graphing-calculator-active

beginning in 1995. We will also go to a new test format, splitting the multiple choice section into two 45-minute parts, one with 25 questions to be taken with *no* calculator, the other with 15 questions, some requiring a graphing calculator. The free response section will remain the same, 6 questions in 90 minutes, and will be designed with graphing calculators in mind. The *extent* to which calculators will be required will be slight at first, but may increase over time as emphases in the curriculum change.

For the next year or so, the Committee will be buried under test development details resulting from our decision. After that we will emerge to begin a careful evaluation of the calculus curriculum, from a to z. Or, if you like, from ϵ to δ . I recommend, however, that every teacher in the AP program get started without us, just as you started without us with graphing calculators. Don't worry about whether you are doing the same thing in your classroom as I am doing in mine; we have all worried about that for too long. Just worry about whether your students are learning calculus and enjoying it. If that's the *only* thing that our classrooms have in common, then our students are still much better off than when we had *everything else* in common at the expense of that.

In the next few years, the vibrations from all these education reform movements will reach a crescendo, and from it all will emerge a new paradigm for teaching and learning. I have seen enough to realize that it will not be confined only to mathematics; nonetheless, it is apparent to any observer that mathematics is leading the way. If Newton does come back to visit us, I hope he waits a few years so that all of this will have had a chance to develop. Then, when he asks to see the Calculus for the Twenty-first Century, we can show him something which Newton would truly appreciate: an entire Renaissance Curriculum.

And I'll bet we get to show it to him on a CD—hooked up to our graphing calculator.

The Baylor School
Box 1337
Chattanooga, TN 37401
DKennedy@UTCVM.Edu

An analyst, surname of Nero
Is my mathematical hero.
Says he, "When in doubt,
I always start out,
'Given $\epsilon > 0 \dots$.'"

James R. Martino
Department of Mathematics
The Johns Hopkins University
Baltimore, MD 21218

Three Problems in Search of a Measure

Jonathan L. King

§P. PREFACE. What do the following three problems have in common?

Poncelet's Theorem. The lefthand figure below shows a pair of ellipses **C** and **E** which happen to have a *circuminscribed polygon*; a polygon which is simultaneously inscribed in the outer ellipse and circumscribed about the inner ellipse.

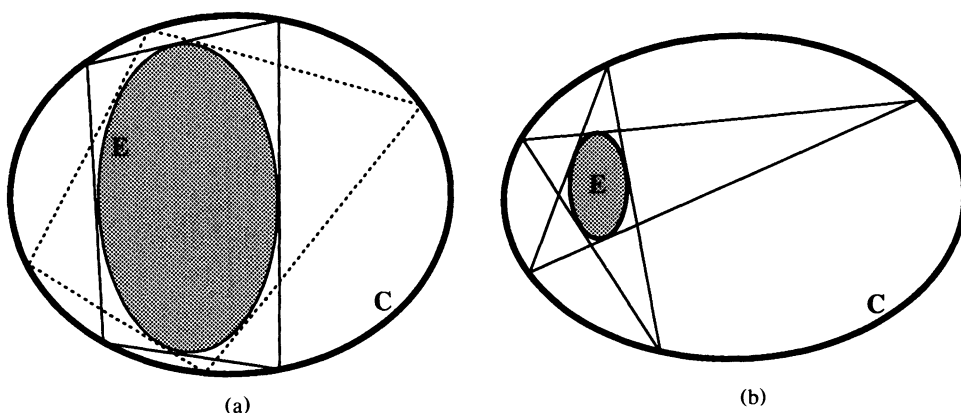


Figure P.1. On the left are two non-degenerate ellipses, with a quadrilateral (solid line) inscribed in **C** and circumscribed about **E**. The dotted-line shows another circuminscribed quadrilateral, which can be thought of as the first quadrilateral “rotated” about **E**. The right hand figure shows a pair of ellipses with a (self-intersecting) circuminscribed pentagon.

Jean-Victor Poncelet's famous CLOSURE THEOREM, published in his *Traité des propriétés projectives* of 1822, asserts that if there exists one circuminscribed n -gon then *any* point on the boundary of **C** is the vertex of some circuminscribed n -gon. Indeed, if we allow the n -gon to continuously change its shape, a circuminscribed n -gon can be continuously rotated around **E**.

How might this theorem be generalized to the case where the **CE** pair has no circuminscribed polygon?

Tarski's Plank Problem. Consider a circular table with diameter 9 feet. At your disposal are many planks, each 1 foot wide and longer than 9 feet. What is the minimum number of planks required to cover the surface of the table? Nine parallel planks certainly cover the tabletop. But can you take fewer and criss-cross them to cover the tabletop? If countably many planks of different widths are permitted, is there a cover using less total width than the parallel one?

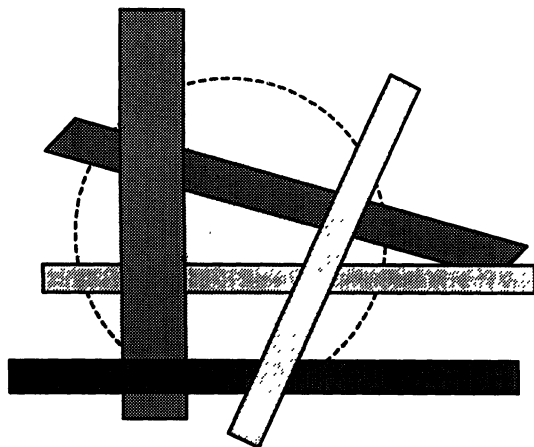


Figure P.2. Thin planks partially covering a round tabletop, in criss-cross fashion. The planks can be thought of as being infinitely long.

Gelfand’s Question. In the table below, row n has the leftmost (high-order) digit of the numbers $2^n, 3^n, \dots, 9^n$, when written in base ten. The “7” in the third row, column 9^n , is the 7 from $729 = 9^3$.

TABLE P.3 The leftmost digits of powers.

n :	2^n	3^n	4^n	5^n	6^n	7^n	8^n	9^n
1:	2	3	4	5	6	7	8	9
2:	4	9	1	2	3	4	6	8
3:	8	2	6	1	2	3	5	7
4:	1	8	2	6	1	2	4	6
5:	3	2	1	3	7	1	3	5
6:	6	7	4	1	4	1	2	5
7:	1	2	1	7	2	8	2	4
⋮				⋯				
99:	6	1	4	1	1	4	2	2
⋮				⋯				

Will a “9” ever occur in the 2-column? Will the row “23456789” appear again? If so, will the set of n such that $\text{row}_n = \text{“23456789”}$ have a *frequency*? —and if it does, will it be rational or irrational? Will a row of all-the-same-digit occur? Will the decimal expansion for an 8-digit prime every appear?

Philosophy. At first glance the three problems seem to have little in common; Poncelet’s theorem is a question about conic sections, the Plank problem about geometric set-inclusion, and Gelfand’s question is number theoretic. Yet it turns out that two of these three are secretly isomorphic.

More significantly, they have a less precise but deeper commonality in that each of the three problems has, sitting somewhere inside it, a natural measure—a measure which, on some collection of sets, is preserved under some family of motions. This basic tool of an **invariant measure** is the theme of this article. It will

turn out that the associated invariant measures make the hidden isomorphism conspicuous.

Anatomy. The next three sections construct in turn natural invariant measures for Poncelet’s Theorem, the Plank Problem and Gelfand’s Question, and can be read essentially independently. All three measures are finite measures which can be interpreted as lengths, areas or volumes. Consequently, the reader need only have an intuitive feel for the size of a set; no theorems of formal measure theory[†] are used. The hope is that the article can be read by motivated undergraduates who are comfortable with integration and elementary topology.

The APPENDIX contains a brief history of each problem or a pointer to such. It alludes to the connection with ergodic theory, and ends with an open problem.

Idiosyncrasy. Use “ $a := b$ ” to mean “ a is defined to be b ”. We use the usual symbol, “ B^c ”, for the complement of a set B ; let “ $A \subset B$ ” denote the set-difference $A \cap B^c$. Symbol $\sqcup_1^\infty B_k$ indicates that the sets $\{B_k\}_k$ in the union happen to be disjoint.

If λ is a measure such as “length” on a space Y , then a “ λ -nullset” $B \subset Y$ has zero length, $\lambda(B) = 0$. An example is a set B consisting of finitely many points. The pair (Y, λ) is called a measure space. A map $\varphi: X \rightarrow Y$ between two measure spaces (X, μ) and (Y, λ) is *measure-preserving* if

$$\mu(\varphi^{-1}B) = \lambda(B), \text{ for all subsets } B \text{ of } Y.$$

A measure-preserving map $R: X \rightarrow X$ from a space to itself is called a *transformation*, which we write in full as $(R: X, \mu)$. Finally, for a point $z \in X$ let $R^n(z)$ denote the n -fold composition $R(R(\dots R(z)\dots))$. The sequence of points $z, R(z), R^2(z), \dots$ is the *orbit* of z under R .

§1 PONCELET’S THEOREM. From any point outside of E , a “righthand tangent to E ” can be consistently chosen. In particular, E gives rise to a “righthand homeomorphism,” R , mapping C to itself.

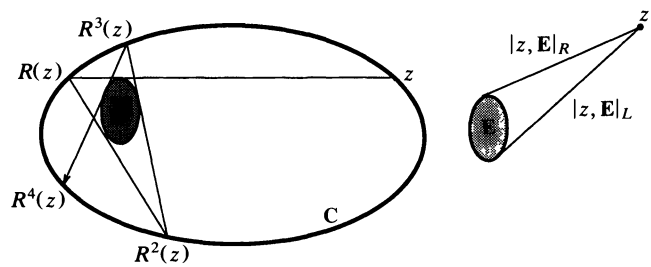


Figure 1.1. Two ellipses, with E properly inside of C . An observer standing at z peering inward at E sees a “righthand” and a “lefthand” tangent to E . Call the righthand tangent map $R: C \rightarrow C$. NOTATION: From a point z outside E , let $|z, E|_L$ and $|z, E|_R$ denote the distance from z to the lefthand and righthand tangent points on E . In the case E is a circle, ie. $|z, E|_L = |z, E|_R$, agree to write $|z, E|$ for the common value.

[†]Originally, this article contained a fourth problem, invented by David Feldman, concerning billiards in the cusp between the curves $y = 1/x$ and $y = -1/x$. Because its solution does use elementary measure theory —and on an infinite measure space— as well as the idea of recurrence in a dynamical system, the discussion of that problem will appear separately, in [KING].

Viewed this way, Poncelet's theorem becomes a statement about the dynamics of the map R and explains why it is called a **CLOSURE THEOREM**; if the orbit of one point z "closes up" in n steps, that $R^n(z) = z$, then the orbit of any point closes up and also in n steps.

In closing up, if the orbit of z "winds around E " p times ($p = 1$ and $n = 4$ in Figure P.1(a); in P.1(b), $p = 2$ and $n = 5$) then every point's orbit would have to wind p times before closing up, since R preserves order along C . This suggests that —after a suitable change of coordinates— R is simply a rigid rotation of a circle by rational rotation number p/n . Let $\mathbb{K} := [0, 1)$ be the half-open interval topologized as a circle, and let \oplus and \ominus denote addition and subtraction modulo 1. For a *rotation number* $\alpha \in \mathbb{R}$, let

$$\rho_\alpha: \mathbb{K} \rightarrow \mathbb{K}: x \mapsto x \oplus \alpha$$

be the corresponding rigid rotation. A "change of coordinates" would then be a homeomorphism $\varphi: C \rightarrow \mathbb{K}$ with a commutative diagram

$$\begin{array}{ccc} C & \xrightarrow{R} & C \\ \downarrow \varphi & & \downarrow \varphi \\ \mathbb{K} & \xrightarrow{\rho_\alpha} & \mathbb{K} \end{array} \quad (1.2)$$

Such a φ satisfying $\varphi \circ R = \rho_\alpha \circ \varphi$ is called a *topological conjugacy* from R to ρ_α . Poncelet's theorem would thus follow from

Lemma 1.3. *For ellipses C , E and mapping R as in Figure 1.1, there is a rotation number $\alpha \in [0, 1)$ and topological conjugacy φ carrying R to ρ_α .*

Indeed, this lemma asserts something even if the R -orbit of z fails to close up—the case when α is irrational.

The appearance of a measure. Arclength measure λ is *invariant* under the rotation ρ_α , that is,

$$\lambda(\rho_\alpha^{-1}(B)) = \lambda(B)$$

for any set B included in \mathbb{K} . We normalize λ to a probability measure, $\lambda(\mathbb{K}) = 1$. A conjugacy φ would lift λ to an R -invariant measure μ on C ,

$$\mu(A) := \lambda(\varphi(A)), \quad (1.4)$$

which we will call *good*: finite, non-atomic (any individual point has zero μ -length) and giving positive length to open intervals.

It is the converse which will help us out:

Any R -invariant good μ gives rise to a topological conjugacy.

In order to define this conjugacy, for points $z, y \in C$ let $[z, y)$ denote the half-open interval on C going counterclockwise from z to y . Next, normalize μ so that $\mu(C) = \lambda(\mathbb{K}) = 1$; now, for any three points z, y, x on C

$$\mu([z, y)) \oplus \mu([y, x)) = \mu([z, x)).$$

Fix any particular point $z_0 \in \mathbf{C}$. Define φ and α by

$$\varphi(y) := \mu([z_0, y)) \quad \text{and} \quad \alpha := \mu([z_0, Rz_0)). \quad (1.5)$$

Thus φ sends z_0 to 0 in \mathbb{K} and is well-defined because of the normalization. Non-atomicity implies continuity of φ and the “positive length” condition insures that φ is invertible. Invariance yields that for any x ,

$$\begin{aligned} \mu([x, Rx)) &= \mu([Rz_0, Rx)) \ominus \mu([Rz_0, x)) \\ &= \mu([z_0, x)) \ominus \mu([Rz_0, x)) \\ &= \mu([z_0, Rz_0)) = \alpha. \end{aligned}$$

In other words, the rotation number α did not truly depend on the arbitrary point z_0 , but only on the R -invariant measure μ . With this, it is easy to verify that φ carries R to ρ_α , as in (1.2).

Our previous lemma can be restated now like this.

Lemma 1.3'. *For ellipses \mathbf{C} and \mathbf{E} , with R as in Figure 1.1: There exists an R -invariant good measure μ .*

When \mathbf{C} and \mathbf{E} are concentric circles, arclength measure along \mathbf{C} is R -invariant. However, the case where they are non-concentric circles seems not as apparent, and the general elliptical case is less evident still. Happily, we can at least assume that the outer ellipse \mathbf{C} is a circle—since any linear map of the plane carries ellipses to ellipses and tangent chords to tangent chords and so carries Figure 1.1 to another just like it. So before even starting to construct μ , one can linearly compress the figure along the major axis of \mathbf{C} to arrange that now \mathbf{C} is a circle.

Rolling the chord. The righthand map, R , generally stretches or shrinks sub-intervals $I \subset \mathbf{C}$, thus making the standard arclength non-invariant. A direct approach to making an R -invariant “length” μ , is to compensate for stretch/shrink by integrating against arclength an appropriately chosen “height function” h whose height varies so as to cancel out the distortion introduced by R . Then the “ μ -length” of a set A would be

$$\mu(A) = \mu_h(A) := \int_A h(z) dz \quad \text{where “} dz \text{” denotes arc-length measure on } \mathbf{C}. \quad (1.6)$$

If h is a continuous function from \mathbf{C} to the positive reals, then automatically μ will be non-atomic and give positive length to open intervals. The issue becomes: *What property does $h(\cdot)$ need to fulfill for the measure $\mu = \mu_h$ to be R -invariant near a point z ?*

In Figure 1.7, the ratio $((z\text{-arc})/\Delta z) \rightarrow 1$ as the angle $\theta \searrow 0$, and similarly $((y\text{-arc})/\Delta y) \rightarrow 1$. Consequently, the “infinitesimal ratio” of the y -arc to the z -arc is

$$\lim_{\theta \searrow 0} \frac{y\text{-arc}}{z\text{-arc}} = \lim_{\theta \searrow 0} \frac{\Delta y}{\Delta z} = \frac{\ell_y}{\ell_z};$$

this last equality comes from the equality $\angle Pyz = \angle Pzy$ of the base angles of isosceles triangle yPz at the right of Figure 1.7.

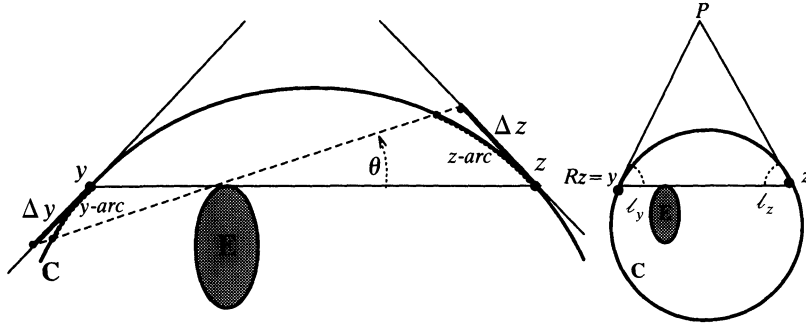


Figure 1.7. Circle C surrounds ellipse E . Let z and y be points at opposite ends of a chord tangent to E , and let ℓ_z and ℓ_y denote their distance to the common point of tangency. Roll the chord through a small angle θ . Its “ z end” sweeps out an arc on C as does its “ y end”; on the corresponding tangent lines, distances Δz and Δy are swept out.

Infinitesimally, the μ -length of the “ z -arc” is simply its arclength times $h(z)$. So for μ to be R -invariant at z , the function h must satisfy that—infinitesimally—the product $h(z) \text{Length}(z\text{-arc})$ equals $h(y) \text{Length}(y\text{-arc})$. In consequence, the invariance condition required of $h(\cdot)$ is

$$h(z) \cdot |z, E|_R = h(y) \cdot |y, E|_L, \text{ where } y = Rz. \quad (1.8a)$$

This uses the previous displayed-equation as well as $\ell_z = |z, E|_R$ and $\ell_y = |y, E|_L$.

Proof of Lemma 1.3': If our inside ellipse E happens to be a circle, then the lefthand and righthand tangential-distances equal a common function $z \mapsto |z, E|$. Then

$$h(z) := 1/|z, E|$$

satisfies (1.8a), and the corresponding μ is the desired R -invariant measure.

To handle the non-circular case, choose some linear map \mathcal{A} which transforms E into a circle, as shown in Figure 1.9. Since a linear map preserves the ratio of lengths of parallel line-segments, we have that

$$\frac{|y, E|_L}{|z, E|_R} = \frac{|\mathcal{A}y, \mathcal{A}E|_L}{|\mathcal{A}z, \mathcal{A}E|_R}.$$

But since $\mathcal{A}E$ is a circle, the subscripts on the second ratio are unnecessary and it can be written $|\mathcal{A}y, \mathcal{A}E|/|\mathcal{A}z, \mathcal{A}E|$. Just as before, then,

$$h(z) := \frac{1}{|\mathcal{A}z, \mathcal{A}E|}, \text{ for all } z \in C, \quad (1.8b)$$

satisfies (1.8a) and makes the measure μ_h —which we will call *Poncelet-measure*—invariant under R .

Exercise: Poncelet-measure for a special case. Suppose C and E are confocal ellipses, with foci at $(0, \pm F)$ and with semi-minor axis lengths of 1 and r

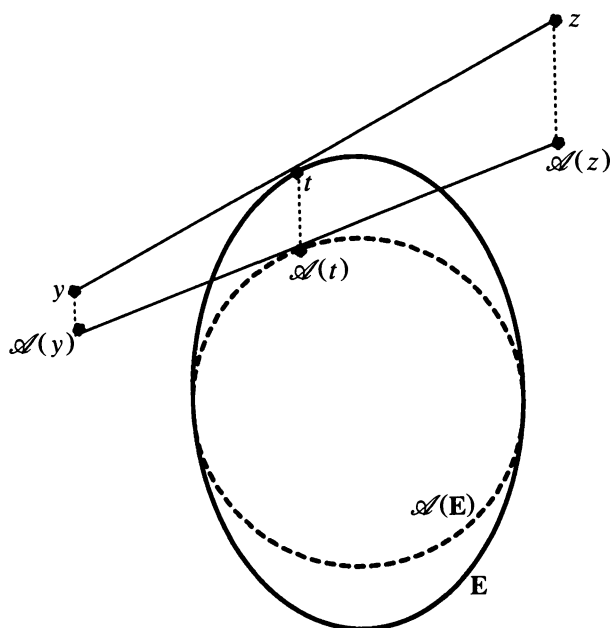


Figure 1.9. Ellipse E is carried by linear map \mathcal{A} to a circle $\mathcal{A}(E)$. For artistic convenience, the linear map illustrated simply compresses the major axis (vertical, in the diagram) of E until equality with E 's minor axis. Each point of chord ytz is carried vertically downward to line-segment $\mathcal{A}(y)\mathcal{A}(t)\mathcal{A}(z)$; here t is the point where the yz -chord is tangent to E .

respectively, ie.

$$\text{C: } x^2 + \frac{y^2}{F^2 + 1^2} = 1^2 \quad \text{and} \quad \text{E: } \frac{x^2}{r^2} + \frac{y^2}{F^2 + r^2} = 1^2,$$

with $0 < r < 1$. Then Poncelet-measure μ_h will be the integral, against arclength measure along ellipse C , of the function

$$h(x, y) = \text{const} \cdot \frac{1}{\sqrt{(1 + x^2 F^2)(r^2 + x^2 F^2)}}, \quad (1.10)$$

where the constant is the square root of $(F^2 + r^2)/(1 - r^2)$. \square

Remark. What can be said when CE has no circumscribed polygon? —that is, when R is conjugate to an irrational rotation. One starts drawing tangent-chords from some point z , as in Figure 1.1, but the polygon-in-progress never closes up. What will be true is that the vertices of this unfulfilled polygon will densely fill out C :

$$\text{Under an irrational rotation } \rho = \rho_\alpha, \text{ the orbit of } 0 \text{ is dense.} \quad (1.11)$$

To establish this, partition the circle \mathbb{K} into M subintervals

$$\left[0, \frac{1}{M}\right), \dots, \left[\frac{M-1}{M}, 1\right).$$

Take $N \geq 1$ smallest such that $\rho^N(0)$ falls inside $[0, (1/M))$. (This certainly will happen for some $N \leq M$ since, by the Pigeonhole Principle, some two of the $M + 1$ points $\{0, \rho(0), \rho^2(0), \dots, \rho^M(0)\}$ fall into the same subinterval.) Thus the ρ^N -orbit of 0 is $(\frac{1}{M})$ -dense. So its ρ -orbit is ε -dense, for all ε . \square

By the way, it is natural to wonder whether the Poncelet rotation-number α is essentially unique; is it independent of linear map \mathcal{A} ? That it is, follows from this challenge:

If rotations ρ_α and ρ_β are topologically conjugate, where $0 \leq \alpha, \beta \leq \frac{1}{2}$, then $\alpha = \beta$.

Another natural question, which is discussed in the appendix, is to wonder

$$\begin{aligned} &\text{Is Poncelet-measure unique? How does} \\ &\mu_h \text{ depend on our choice of linear map } \mathcal{A}? \end{aligned} \tag{1.12}$$

§2 THE PLANK PROBLEM. Experimentation with strips of paper as “planks” tends to suggest that a disk \mathbf{D} cannot be efficiently covered unless the strips are parallel, which leads one to this

Plank Conjecture. *Suppose $(w_n)_{n=1}^\infty$ are the widths of a countable family of planks which cover disk \mathbf{D} . Then*

$$\sum_{n=1}^\infty w_n \geq \text{Width}(\mathbf{D}).$$

If $\sum_{n=1}^\infty w_n$ actually equals $\text{Width}(\mathbf{D})$, the diameter of the disk, then the planks must be parallel to one another.

Measuring area differently. We can take \mathbf{D} to be the closed radius-1 disk centered at the origin and ask that a *plank* P be the closed region between two parallel lines, *both* of which contact \mathbf{D} . The *width* of P is the perpendicular distance between these parallel edges.

In order to cover the disk by small total width, one is tempted to use planks which pass near the center of \mathbf{D} , since such planks P cover more than their fair share of the area of the disk; more than $\text{Width}(P)/\text{Width}(\mathbf{D})$. Yet if many planks pass near the center, they must waste some of their area by overlapping one another. Conversely, while it is easy to arrange that planks passing near the boundary of \mathbf{D} be disjoint, such planks cover less than their fair share.

An answer to the Plank Problem is not obvious because the area covered by a plank P is not invariant under moving P over the disk. An analogy with Poncelet’s theorem is useful. The only “obvious” case where a circumscribed polygon can be rotated about \mathbf{E} , is when the two ellipses are concentric circles. And this is exactly the case where normal arclength on \mathbf{C} is evidently invariant under the transformation R . What saved the day, in the case of general ellipses, was the discovery of a different way of measuring length on \mathbf{C} , the measure μ_h of (1.6), which indeed is invariant under motion by R .

Returning to our planks, the analogy suggests looking for a new way to measure area on \mathbf{D} so that this “new area” —at least for planks— is invariant under rigid motions. We might expect to construct this measure, ν , as we did with Poncelet-measure, by integration of some nice positive height function $h: \mathbf{D} \rightarrow \mathbb{R}_+$ against area. If we build ν in this way, then any subset of the disk with zero ν -area would have to have zero area, as well. Also, the desired invariance of ν under rigid motions would mean there is a positive constant \hbar so that

$$\nu(P) = \hbar \cdot \text{Width}(P), \quad \text{for any plank } P. \tag{2.1}$$

One clarification is in order. Since ν is to be a measure on \mathbf{D} , we need to interpret a plank P as subset of the disk and agree to regard “ P ” as the set $P \cap \mathbf{D}$. With this convention, \mathbf{D} itself is a plank.

Assuming now that someone with outstanding geometric intuition has provided us with such a ν , let's use it to verify the conjecture.

Provisional Proof of the Plank Conjecture. If planks $\{P_n\}_1^\infty$ cover \mathbf{D} then

$$\begin{aligned} \text{Width}(\mathbf{D}) &= \frac{1}{h} \nu(\mathbf{D}) = \frac{1}{h} \nu\left(\bigcup_n P_n\right) \\ &\leq \frac{1}{h} \sum_n \nu(P_n) = \sum_n \text{Width}(P_n). \end{aligned} \quad (2.2)$$

This shows that no cover can use less total width than a parallel cover.

To show that *only* parallel covers use minimal width, suppose now we have a cover realizing equality in (2.2), that is, $\sum_n \nu(P_n) = \nu(\bigcup_n P_n)$. Consequently, the intersection $P_i \cap P_j$ of any pair of planks must have no ν -area and thus no area. This means that in a minimal cover any two planks have disjoint interiors; no two planks *overlap*. So we need but prove that

Any two planks P and R of a non-overlapping cover, are parallel.

To see this, consider any plank Q of the cover which intersects segment L_1 in the figure below.

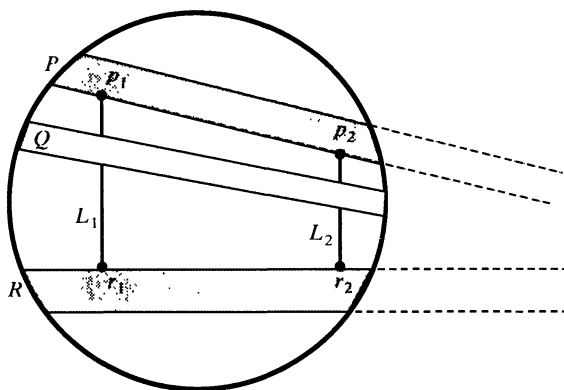


Figure 2.3. Given two planks P and R , choose distinct parallel line-segments $L_1 = \overline{p_1 r_1}$ and $L_2 = \overline{p_2 r_2}$; the L_i are “open” segments, not containing their endpoints. Endpoints p_i and r_i are on neighboring edges of P and R , respectively. Since L_1 and L_2 are disjoint from P and R , they must be covered by the other planks.

We can have chosen the point p_1 to be in the interior of \mathbf{D} ; so Q cannot have L_1 as an edge. Therefore, since Q overlaps neither P nor R it must run *between* their neighboring edges; hence Q crosses both L_1 and L_2 . Since the L_i are parallel, intervals $Q \cap L_1$ and $Q \cap L_2$ have the same length. Consequently,

$$\sum_{Q \in \mathcal{Q}_1} \text{Length}(Q \cap L_1) = \sum_{Q \in \mathcal{Q}_1} \text{Length}(Q \cap L_2),$$

where \mathcal{Q}_1 consists of all planks in the cover which overlap L_1 . But these planks do not overlap each other. Consequently the closed intervals $\{Q \cap L_i\}_{Q \in \mathcal{Q}_1}$ are non-overlapping, and we may conclude that $\text{Length}(L_1) \leq \text{Length}(L_2)$. Reversing the roles of L_1 and L_2 shows that they have equal length. And this means that P and R were parallel all along. \square

Both halves of the proof, the inequality and the parallelness, hinge upon our not-yet-known-to-exist plank-measure ν . The rotational symmetry of condition (2.1) suggests looking at rotationally symmetric geometric figures as a possible source of plank-measure. It turns out that the geometric insight which will provide plank-measure was made by Archimedes in his treatise *On the Sphere and the Cylinder*.

Archimedes’ area-preserving map of the Earth. The figure below shows how to project the globe onto a cylinder of paper in order to get a map of the Earth on which countries of equal spherical-area are represented by map-regions of equal (planar) area.

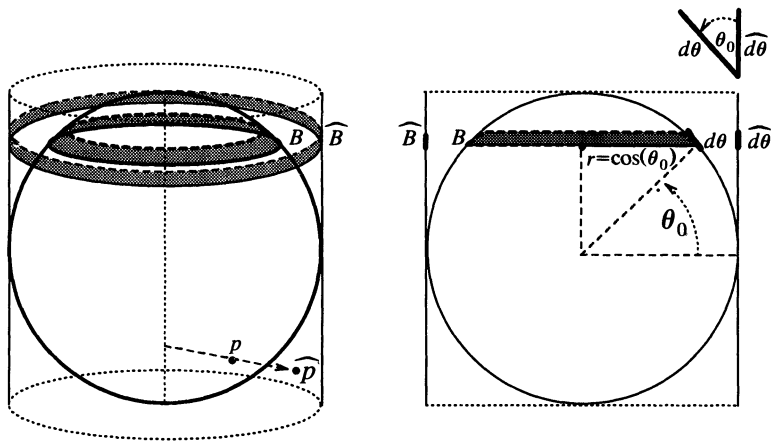


Figure 2.4. A sphere S and a (finite) cylinder C are tangent along their common equator. The height of C equals the sphere’s diameter. For a point $p \in S$, let $\hat{p} \in C$ be its radial-projection: A ray starting from and perpendicular to the axis-of-symmetry of C passes through p and intersects the cylinder at \hat{p} .

Archimedes proved that for any region B on the sphere,

$$\text{Area}_C(\widehat{B}) = \text{Area}_S(B).$$

Since both the sphere and the cylinder are rotationally symmetric, it suffices to establish this when B is a band. Figure 2.4 shows, drawn on a sphere of radius 1, a band B with lower edge at latitude θ_0 , and upper edge infinitesimally higher at latitude $\theta_0 + d\theta$. Since the radius of the bottom edge of B is $r = \cos(\theta_0)$,

$$\text{Infinitesimal Area}_S(B) = (2\pi r) \cdot d\theta = 2\pi \cos(\theta_0) d\theta.$$

The projected band \widehat{B} has radius 1 and infinitesimal width $\widehat{d\theta} = d\theta \cdot \cos(\theta_0)$, by similar triangles. Consequently

$$\text{Infinitesimal Area}_C(\widehat{B}) = (2\pi \cdot 1) \cdot \widehat{d\theta} = 2\pi d\theta \cos(\theta_0).$$

The two infinitesimal areas are thus seen to be equal. □

Constructing plank measure. Area-measure of a sphere whose equatorial-plane is \mathbf{D} , when projected down to \mathbf{D} , has the desired plank property. Define ν by

$$\nu(A) := \text{Area}_S(A_S).$$

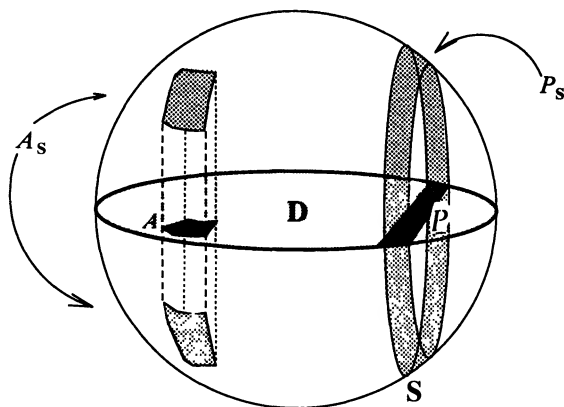


Figure 2.5. A sphere S whose equatorial-plane is our disk D . For any subset $A \subset D$, let A_S be its orthogonal projection on the sphere. Thus the projection of a subdisk concentric with D would be a pair of “polar icecaps”.

The image of a plank $P \subset D$ is a band P_S on the sphere. The sphere can then be oriented so that the band projects on the cylinder to a cylindrical-band \widehat{P}_S whose width necessarily equals the width of the original plank P . Thus

$$\nu(P) = \text{Area}_S(P_S) = \text{Area}_C(\widehat{P}_S) = 2\pi \cdot \text{Width}(P),$$

which demonstrates that every plank’s measure is a *constant* times its width. Moreover, this establishes that the natural value of “Plank’s Constant”, \hbar , is 2π .

§3 GELFAND’S QUESTION. Questions about high-order digits of powers, Table P.3, provide a perfect setting for students to formulate and test numerical conjectures. Here are some “mysterious empirical facts” which can be found by judicious use of a computer. For a positive x , let $\langle\langle x \rangle\rangle$ denote the high-order non-zero digit of the base-ten expansion of x . Thus $\langle\langle \text{Plank’s Constant} \rangle\rangle = 6$ and $\left\langle\left\langle \frac{e}{1000} \right\rangle\right\rangle$ equals 2.

Mysterious observations? Studying the 2-column of Table P.3,

2 4 8 1 3 6 1 2 5 1 2 4 8 1 3 6 1 2 5 ... $\langle\langle 2^n \rangle\rangle$...

one discovers that the nine digits *do* appear to have frequencies[‡]—and that they are unequal. Indeed, letting $\text{fr}(d)$ denote the *frequency* of digit d ,

$$\text{fr}(d) := \lim_{N \rightarrow \infty} \frac{1}{N} |\{n \mid 1 \leq n \leq N \text{ and } \langle\langle 2^n \rangle\rangle = d\}|,$$

the computer suggests that the frequencies decrease steadily as d increases from 1 to 9:

$$\text{fr}(1) \approx .301, \quad \text{fr}(2) \approx .176, \quad \dots \quad \text{fr}(9) \approx .045.$$

The first empirical surprise is that digit d occurs with the same frequency $\text{fr}(d)$ in *every* column of the table. This, despite the fact that the columns are decidedly

[‡]The existence of frequencies is not a foregone conclusion. In the sequence of high-order digits $(\langle\langle n \rangle\rangle)_{n=1}^\infty$ of *all* natural numbers, the digit “1” does not have a frequency, since its upper-density is five times its lower-density.

different—they all have patterns which appear in no other column. For example, the sequence “248136” occurs infinitely often in column 2, yet will never be witnessed by the other columns. Indeed, *no* sequence of length six which appears in a column, *ever* appears in any other column.

Every digit occurs in every row of Table P.3, and yet an all-the-same row never appears. This suggests looking at *joint frequencies*. Let $F_{2:6:9}(d, d', d'')$ be the frequency of rows n which have d in column 2, digit d' in column 6 and d'' in column 9. Some pairs of columns seem to be independent —columns 2:3 for instance— in that there is approximate equality

$$F_{2:3}(d, d') \approx \text{fr}(d)\text{fr}(d').$$

Yet columns 4 and 5 appear strongly non-independent. Only one digit d ever occurs simultaneously in columns 2:5. Columns 2:4 don't have independence, yet columns 3:6 do. But 3:9 do not. And so on...

The real startler comes, however, when you vary the “seed.” Row_{n+1} is obtained by multiplying row_n by the tuple $(2, 3, \dots, 9)$. Table P.3 starts with a “zero-th row” of $(1, 1, \dots, 1)$. But there is no reason to use all the same integer—and upon consideration, why even restrict to integers? We could start with a “zero-th row” (s_2, \dots, s_9) of positive real numbers as *seeds* and then study frequencies; let the symbol $F_{\mathbf{m}:\mathbf{m}}^{s;\mathbf{m}}(d, d')$ mean the frequency of those n for which simultaneously $\langle\langle s\mathbf{m}^n \rangle\rangle = d$ and $\langle\langle s'\mathbf{m}'^n \rangle\rangle = d'$.

At first this seems like a waste of time, since the seed appears to have no effect on individual columns; no matter which of $2, \dots, 9$ one takes for the multiplier \mathbf{m} , apparently $F_{\mathbf{m}}^s(d) = \text{fr}(d)$ independently of the seed s . But this independence abruptly disappears for joint frequencies. An energetic computer will discover, for example, the oddity that the mapping

$$s \mapsto \mathbf{F}_{2:5}^{s, s}(3, 3) \tag{3.1}$$

is continuous—but fails to be differentiable at exactly three points in the interval $(1, 10)$.

Converting to a dynamical system. The mathematics is simpler if we first just analyze the statistics of a single column—the 2-column perhaps or, more generally, the \mathbf{m} -column for some fixed positive multiplier \mathbf{m} . Effectively we are studying the “multiply map”

$$T: \mathbb{R}_+ \rightarrow \mathbb{R}_+ : x \mapsto \mathbf{m}x$$

applied to an initial seed s ; we are following the T -orbit of s , taking the measurement $\langle\langle T^n(s) \rangle\rangle$ along it. Since x and $10x$ have the same first digit, we might as well identify numbers whose ratio is a power of 10. Doing this collapses all these half-open intervals,

$$\dots \left[\frac{1}{10}, 1 \right), [1, 10), [10, 100), \dots$$

together, thus effectively wrapping the positive-reals \mathbb{R}_+ into a circle. Insofar as our measurements are concerned, then, the multiply map is just some homeomorphism of a circle.

The easiest way to identify $10x$ with x is simply to take logarithms base-ten and then discard the integer part. Letting ψ denote this identification and letting

$\mathbb{K} = [0, 1)$ be the circle as before, we get this commutative diagram:

$$\begin{array}{ccc} \mathbb{R}_+ & \xrightarrow{T} & \mathbb{R}_+ \\ \downarrow \psi & & \downarrow \psi \\ \mathbb{K} & \xrightarrow{\rho_\alpha} & \mathbb{K} \end{array} \quad \text{where } \psi(x) := (\log_{10} x)_{\bmod 1} \quad (3.2)$$

Since the logarithm converts multiplication to addition, transformation $\rho_\alpha := \psi \circ T \circ \psi^{-1}$ is simply rotation $x \mapsto x \oplus \alpha$ on the circle by $\alpha := \log_{10}(\mathbf{m})$, where “ \oplus ” means addition modulo 1. Thus the high-order digit of a number x is d if and only if $\psi(x)$ is in the half-open interval

$$I_d := [\log(d), \log(d + 1)).$$

(Here, and henceforth, “log” means \log_{10} .) These nine intervals $\{I_d\}_{d=1}^9$ partition the circle.

An isomorphism. The underlying transformation for Gelfand’s Question in the case of a single column, we’ve discovered, is a rigid rotation of a circle—this is the same transformation found hidden in Poncelet’s theorem. Can this isomorphism be made more explicit? Yes it can.

The ellipses **C** and **E** in Figure 1.1 gave rise to a rotation number α and commutative diagram 1.2. Use the φ of that diagram to lift the intervals $\{I_d\}_1^9$ to a partition $\{J_d\}_1^9$ on **C**, where $J_d := \varphi^{-1}(I_d)$, and let $\mathbf{m} := 10^\alpha$.

Effectively $\varphi^{-1} \circ \psi$ is an isomorphism from the Gelfand \mathbf{m} -system to the Poncelet **CE**-system: Any question about the \mathbf{m} -column with seed s is mapped to the corresponding question about the Poncelet-orbit of $z := \varphi^{-1}(\psi(s))$. This R -orbit of z lands in intervals

$$J_{d_0}, J_{d_1}, J_{d_2}, \dots, J_{d_n}, \dots$$

precisely so that the high-order digit of $s\mathbf{m}^n$ is invariably d_n .

Equidistribution. Now that we know that the **2**-column is isomorphic to the orbit of 0 on the circle under rotation by $\log(2)$, what do we know? Since $\log(2)$ is irrational this orbit is dense, courtesy remark 1.11, and so every digit appears infinitely often in column **2**. We now find ourselves in the embarrassing circumstance of knowing there are integers n where 2^n starts with “9”—without having the foggiest notion of a single one.

However, *denseness*—a property which ignores the *time* when points appear in the orbit—is not sufficient to determine the *frequency* of “9”, or whether it has a frequency at all. A more refined “randomness” property is needed: A sequence z_0, z_1, \dots in the circle is *equidistributed* if

$$\text{For any interval } I \subset \mathbb{K}: \quad \lim_{N \rightarrow \infty} \frac{1}{N} |\{n | 0 \leq n < N \text{ \& } z_n \in I\}| = \text{Length}(I). \quad (3.3)$$

Weyl’s EQUIDISTRIBUTION THEOREM for the circle says that for any irrational α :

$$\begin{array}{l} \text{For any point } z, \text{ the sequence } z, z \oplus \alpha, z \oplus 2\alpha, \dots, z \oplus n\alpha, \dots \\ \text{is equidistributed in } \mathbb{K}. \end{array}$$

For a multiplier $\mathbf{m} \in \{2, \dots, 9, \pi, e, \text{ etc.}\}$ which is not a rational power of ten, the corresponding rotation number $\alpha = \log(\mathbf{m})$ is irrational. Weyl’s theorem explains the mysterious frequencies $\text{fr}(d)$:

$$\text{fr}(d) = \text{Length}(I_d) = \log\left(\frac{d + 1}{d}\right), \quad \text{for } d = 1, 2, \dots, 9.$$

So the frequencies are irrational—in fact, transcendental. Weyl’s theorem also explains why these frequencies do not depend on the seed.

Randomness and Ergodicity. An irrational rotation ρ is “topologically random” in the sense that each orbit is dense. Restated, the orbit of any point $a \in \mathbb{K}$ hits any non-empty open set U : *There exists n so that $a \in \rho^{-n}(U)$.*

An analogue of this, a notion of measure-theoretic “randomness”, is that the orbit of any set A of positive length hits any other positive-length set U : *There exists n so that $\lambda(A \cap \rho^{-n}(U)) > 0$.* This measure-theoretic property makes sense for any measure-preserving transformation $(T : X, \mu)$, and has this equivalent formulation:

Any T -invariant[†] set B is —up to a nullset— either empty or is the whole space. Either B or $X \setminus B$ is a nullset.

The connection between our irrational rotation ρ and ergodicity is twofold. On the one hand, an irrational rotation is ergodic. On the other hand, any ergodic transformation satisfies an abstract equidistribution theorem, due to Birkhoff, called the POINTWISE ERGODIC THEOREM. This ergodic theorem is sufficiently strong to imply both Weyl’s theorem and —a result we will need shortly— the KRONECKER-WEYL THEOREM. A formulation of Birkhoff’s theorem appears in the appendix, as well as a proof that irrational rotations are ergodic.

It is the possibility that a transformation can *fail* to be ergodic which gives rise to the instability of frequencies $F_{\mathbf{m};\mathbf{m}'}^{s,s'}(d,d')$ as the seeds s and s' are varied.

Non-equidistribution. When $\mathbf{m} = 10^{p/q}$ is the multiplier then $\rho = \rho_\alpha$ is a rational rotation. Take, for example, $\alpha = 1/3$. The ρ -orbit of any “seed” $z \in \mathbb{K}$ thus has exactly 3 points. The observed frequency of a digit d does not remain constant as the seed is varied; for example, the frequency that the z -orbit hits I_2 is $\frac{1}{3}$ for any $z \in I_2$, but is zero for $z \in I_3$.

We also notice another failure—that $\rho_{1/3}$ is not ergodic. Given any $A \subset \mathbb{K}$, the set $B := A \cup \rho^{-1}(A) \cup \rho^{-2}(A)$ is ρ -invariant. Taking an A of positive length less than $1/3$ gives a non-trivial invariant set B . This explains lack of constancy for frequencies; the orbit of any seed $z \in B^c$ hits B with zero frequency and yet the measure of B is positive.

Happily, any transformation can be decomposed into disjoint ergodic transformations. In the case of rotation by $1/3$, write the circle as a disjoint union

$$\mathbb{K} = \bigsqcup_{0 \leq c < 1/3} K_c, \quad \text{where } K_c := \{c, c \oplus \tfrac{1}{3}, c \oplus \tfrac{2}{3}\}, \tag{3.4a}$$

of ρ -invariant 3-point sets K_c . For a point z in the circle, the particular $K = K_c$ containing z is called the *ergodic component* of z because, when K is equipped with this ρ -invariant probability measure ν_K ,

$$\nu_K(\{c\}) = \nu_K(\{c \oplus \tfrac{1}{3}\}) = \nu_K(\{c \oplus \tfrac{2}{3}\}) = \tfrac{1}{3}, \tag{3.4b}$$

then the system $(\rho : K, \nu_K)$ is ergodic. As K ranges over the ergodic components of ρ , the measures ν_K form a disintegration of arclength measure λ . This is why (3.4) is called the *ergodic decomposition* of $(\rho : \mathbb{K}, \lambda)$.

Joint frequencies. Frequencies $F_{2,3}(d,d')$ are measurements made on the direct product of two “multiply maps” and so $\psi \times \psi$ (see commutative diagram 3.2)

[†]“ T -invariant” can be taken to mean either $T^{-1}(B) = B$, or the superficially weaker statement that the symmetric difference $B \Delta T^{-1}B$ is a nullset. It is easy to check that if B satisfies the latter invariance, then B can be altered by a nullset to fulfill the stronger invariance.

carries this system to a rotation of the torus $\mathbb{K}^{\times 2} = \mathbb{K} \times \mathbb{K}$,

$$\rho_\alpha \times \rho_\beta : (z, z') \mapsto (z \oplus \alpha, z' \oplus \beta),$$

where $\alpha = \log(2)$ and $\beta = \log(3)$. So $F_{2:3}^{s,s'}(d, d')$ is simply the frequency that the $\rho_\alpha \times \rho_\beta$ -orbit of $(z, z') := (\psi(s), \psi(s'))$ hits rectangle $I_d \times I_{d'}$.

Analogous to the one-dimensional case, if $\rho_\alpha \times \rho_\beta$ is “random” —is ergodic— one might expect this frequency to be the measure of the rectangle, that is, its area $\text{fr}(d)\text{fr}(d')$. This, as in Weyl’s theorem, is a statement of equidistribution and is enunciated in the well-known Kronecker-Weyl theorem.

Real numbers $\{\alpha_1, \dots, \alpha_L\}$ are *rationally independent* if they are linearly independent over \mathbb{Q} ; the only integral solution to $N_1\alpha_1 + \dots + N_L\alpha_L = 0$ is the all-zero $N_1 = 0, \dots, N_L = 0$ solution. On $\mathbb{K}^{\times L}$, the L -dimensional torus of L -dimensional volume 1, a sequence of points x_0, x_1, \dots is *equidistributed* if it hits every sub-block $I_1 \times \dots \times I_L$ of the torus with frequency equal to the subblock’s volume.

Kronecker-Weyl Theorem. *Numbers $1, \alpha_1, \dots, \alpha_L$ are rationally independent if and only if under the action of rotation $\rho_{\alpha_1} \times \dots \times \rho_{\alpha_L}$ on the L -dimensional torus, every orbit is equidistributed. (This is also equivalent to ergodicity of this toral rotation.)*

As an application, since $\{1, \log(2), \log(3)\}$ are rationally independent, we have that

$$F_{2:3}^{s,s'}(d, d') = \log\left(1 + \frac{1}{d}\right)\log\left(1 + \frac{1}{d'}\right),$$

regardless of the seeds s and s' .

What happens with non-independence? Take, for example, $\alpha = \log(2)$ and $\beta = \log(5)$. Here, the equality $2 \cdot 5 = 10$ translates to the non-zero rational relation $\alpha + \beta = 1$, and so $\rho_\alpha \times \rho_\beta$ is not ergodic. To see this, regard the 2-dimensional torus $[0, 1) \times [0, 1)$ as a square. Then the ergodic components are “wrap-around” diagonals running northwest to southeast on this square; each number $c \in [0, 1)$ gives us a diagonal K_c which is $(\rho_\alpha \times \rho_\beta)$ -invariant:

$$\mathbb{K}^{\times 2} = \bigsqcup_{0 \leq c < 1} K_c, \quad \text{where } K_c := \{(x, y) | x \oplus y = c\}.$$

Each diagonal K_c is topologically a circle whose ergodic measure λ_c is simply normalized arclength,

$$\lambda_c(A) := \frac{\text{Length}(A \cap K_c)}{\text{Length}(K_c)}, \quad \text{where } A \subset \mathbb{K}^{\times 2}.$$

In this **2:5** case, the system $(\rho_\alpha \times \rho_\beta : K_c, \lambda_c)$ is isomorphic to irrational rotation on the diagonal “circle” K_c by rotation number α (or, equivalently, β).

Weyl’s theorem then asserts that $F_{2:5}^{s,s'}(d, d')$ is the λ_c -length of the intersection of rectangle $I_d \times I_{d'}$ with the diagonal K_c which contains the ψ -image of (s, s') . This is the diagonal with $c := \psi(s) \oplus \psi(s')$. For example, the **2:5** column-pair of Table P.3 corresponds to $s = s' = 1$ and so $c = 0$; therefore $K = K_c$ is the main diagonal $x + y = 1$. It should be easy to figure out the frequency of seeing a doubled digit (d, d) , no?

Scanning the **2** and **5** columns of Table P.3 shows that “3” occurs doubled, $2^5 = \underline{32}$ and $5^5 = \underline{3125}$. It is not surprising that no other digit occurs doubled since, if $d \neq 3$, then the main diagonal does not pass through the square $I_d \times I_d$ and consequently the frequency of (d, d) is zero. Conversely, if you traverse the

main diagonal K by heading southeast, you will enter the $I_3 \times I_3$ square at the point $(\log(3), 1 - \log(3))$ and exit the square at $(1 - \log(3), \log(3))$. The fraction of your time spent inside the square was

$$F_{2;s}(3, 3) = [1 - \log(3)] - \log(3) = 1 - \log(9),$$

which is irrational. (By the way, this picture of a northwest to southeast diagonal explains the three points of non-differentiability of the function $g(s) := F_{2;s}^s(3, 3)$ mentioned in (3.1). As the seed s increases, the corresponding diagonal moves northeast and the length of it within the $I_3 \times I_3$ square varies linearly *except* when the diagonal passes over a corner of $I_3 \times I_3$.)

The general case. What are the ergodic components of toral rotation $\rho_{\alpha_1} \times \cdots \times \rho_{\alpha_L}$?—it being clear now that to compute joint-frequencies of the form $F_{\alpha_1; \dots; \alpha_L}^{z_1; \dots; z_L}$, one needs a geometric description of these ergodic components. The Kronecker-Weyl theorem tells us how, in principle, to do this. However, computing frequencies for a specific set of α_i will require a bit of geometry.

Consider a maximal subset of $\{\alpha_1, \dots, \alpha_L\}$ which, were “1” adjoined, would comprise a rationally independent collection; suppose it has $M \geq 1$ members. Let G be the closure of the orbit of $(0, \dots, 0)$; in other words, the closure in the L -dimensional torus of tuples

$$((n\alpha_1)_{\text{mod } 1}, (n\alpha_2)_{\text{mod } 1}, \dots, (n\alpha_L)_{\text{mod } 1})$$

as n ranges over the integers. This G will be an M -dimensional subtorus (indeed, a subgroup of $\mathbb{K}^{\times L}$). So the ergodic components of $\rho_{\alpha_1} \times \cdots \times \rho_{\alpha_L}$ are merely the translates of this subtorus. Consequently, the frequency

$$F_{\alpha_1; \dots; \alpha_L}^{z_1; \dots; z_L}(d_1, \dots, d_L)$$

is the M -dimensional “area” of the cross-section of the block (rectangular parallelepiped) $I_{d_1} \times \cdots \times I_{d_L}$ which is sliced by the particular subtorus that contains the point (z_1, \dots, z_L) . This is the subtorus $G \oplus (z_1, \dots, z_L)$.

Even for the case of three rotation numbers with exactly two being rationally independent, determining frequencies might be tricky. Here one wants to compute the area of the region which is the intersection of a plane slicing through a 3-dimensional block. Depending on the tilt of this plane, the intersection could be a triangle, a quadrilateral, a pentagon or a hexagon.

So we end up by finding another point of commonality among the problems of Poncelet, Tarski and Gelfand: Existence of a natural invariant measure does not, alas, imply that the measure is going to be easy to compute...

§A APPENDIX. Here we give a terse history of the three questions, as well some related conundrums. A technical note: In this article all measures were tacitly Borel measures, and all sets and functions were Borel measurable.

Poncelet’s theorem, history. Jean-Victor Poncelet was an officer of engineers in Napoleon’s army during the invasion of Russia. Like other invasions of Russia this one failed, and on November eighteenth of 1812, Poncelet was captured in the retreat from Moscow. During his 1812–1814 captivity, which he later wrote about in PONCELET 1862, he developed the notions of Projective Geometry which led him to the Closure theorem. (For a fuller history, see BELL 1937.)

Commentary. The idea of using an invariant to prove Poncelet’s theorem goes back to Jacobi and Bertrand, according to SCHOENBERG 1983; and Schoenberg’s

beautiful paper gives an efficient proof using an *invariant integral*—which in this context is the same thing as an invariant measure.

Nonetheless, Schoenberg's proof uses non-elementary notions: The Brouwer fixed-point theorem and facts in projective geometry concerning the polar line of a point. In contrast, by focusing on an invariant *measure* for Poncelet's transformation, even the small amount of computation in Schoenberg is entirely avoided. This is merely a change in viewpoint, not a change in proof, yet it has the advantage of yielding a theorem even in the case that the given pair of ellipses have no circumscribed polygon, and thereby suggests the link with Gelfand's question. By avoiding Brouwer's theorem, which is non-constructive in nature, one can actually compute the invariant density, eg. (1.10), and thus in principle determine whether the rotation number is rational so as to ascertain whether a circumscribed polygon exists.

An extensive history and “pre-history” of Poncelet's theorem, as well as several proofs, appear in a paper by BOS, KERS, OORT, RAVEN 1987, a paper which I came upon after this article was written.

There is a connection between the Poncelet-measure determined by confocal ellipses C and E , and the invariant “billiard measure” of an elliptical billiard table. The ergodic components of the billiard measure of C turn out to be the 1-parameter family of Poncelet CE -measures obtained by varying E , the inner confocal ellipse. Billiard measure is defined in a companion article to this one, [KING], where it is used to study a billiard question on a table with a cusp at infinity.

Plank Problem, history. Archimedes proved in proposition 3 of Book II of *On the Sphere and the Cylinder*, see HEATH 1897, that a plane orthogonal to a diameter of a sphere cuts the sphere into two regions with areas in the same ratio as the lengths of the two pieces of the diameter. Since the areas of the sphere and cylinder were known to Archimedes, this proposition is tantamount to showing that radial projection is area-preserving. I am indebted to C. E. Thompson for indicating this proposition to me.

Given here was the elementary case of *Tarski's Plank Problem*, proposed in 1932. What Tarski had asked was this:

Suppose a convex compact region D in the plane is covered by countably many planks. Must the sum of their widths dominate the width of D ?

(The *width* of a shape is the width of the narrowest single plank which covers it.) This problem does not immediately follow from the disk case, since some convex shapes —the equilateral triangle being a notorious example— have a width exceeding the diameter of its inscribed circle. Almost 20 years later, Tarski's problem was solved by an ingenious argument of THØGER BANG, 1951. The elementary version of Tarski's problem suggests the following question which, to my knowledge, is open.

Question A.1. What is the infimum, $W(r)$, of total widths $\sum_{n=1}^{\infty} \text{Width}(P_n)$ taken over all systems $\{P_n\}_1^{\infty}$ of planks which cover the annulus of outer-radius 1 and inner-radius r ?

Gelfand's Question, history. The “frequencies of $\langle\langle 2^n \rangle\rangle$ ” problem appears on page 37 of AVEZ 1966, where it is attributed to Gelfand.

Weyl's theorem is usually proved by noting that definition (3.3) of equidistribution is equivalent to this formulation: The sequence x_0, x_1, \dots is *equidistributed* in

the circle \mathbb{K} if

$$\text{For any continuous } g: \mathbb{K} \rightarrow \mathbb{C}: \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} g(x_n) = \int g d\lambda \quad (\text{A.2a})$$

Weyl's theorem concerns those g for which, for each z , the sequence $\{z \oplus n\alpha\}_{n=0}^{\infty}$ is equidistributed. Since the set of such g forms a closed subspace (in the supremum-norm) of the space of all continuous functions, the Stone-Weierstrass theorem tells us that it is enough to check this when g is a group-character, where it is easily verified.

The Pointwise Ergodic Theorem. There is a more general “equidistribution theorem” which applies when T is a measure-preserving transformation on probability space (X, μ) . In the special case when T is ergodic, Birkhoff's theorem states that for any $g \in L^1(\mu)$ the limit and equality below exist.

$$\text{For a.e. } x \in X: \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} g(T^n x) = \int g d\mu. \quad (\text{A.2b})$$

(The nullset of “bad” x violating equality will generally depend on g ; there is no universal full-measure set working for all functions. An earlier ergodic theorem, convergence in the L^2 -norm, was proven by von Neumann.) The pointwise ergodic theorem is a far-reaching generalization of the Strong Law of Large Numbers (SLLN). It also quickly implies Weyl's theorem —that continuous functions g have *no* bad points— once irrational rotations are shown to be ergodic. We show that now.

Proposition A.3. *Any irrational rotation $\rho = \rho_\alpha$ is ergodic.*

Proof: Suppose ρ -invariant set B has positive mass. Take a short open interval I in the circle such that

$$\lambda(I \cap B) \geq (.99)\lambda(I).$$

By the density of the orbit of an endpoint of I , we can find times k_1, \dots, k_M such that the translated intervals $\rho^{k_m}(I)$ are disjoint, and

$$\lambda\left(\bigsqcup_{m=1}^M \rho^{k_m}(I)\right) \geq .98.$$

But $B = \bigcup_{m=1}^M \rho^{k_m}(B) \supset \bigsqcup_{m=1}^M \rho^{k_m}(I \cap B)$. Consequently

$$\begin{aligned} \lambda(B) &\geq \sum_{m=1}^M \lambda(\rho^{k_m}(I \cap B)) \\ &\geq \sum_{m=1}^M (.99)\lambda(\rho^{k_m}(I)) \geq (.99)(.98) > .97. \end{aligned}$$

Of course this “.97” could have been made as close to 1 as desired, so $\lambda(B^c) = 0$. \square

The upshot is that Birkhoff's theorem simultaneously contains the probabilistic equidistribution of SLLN and the topological equidistribution of Weyl.

Poncellet-measure is unique. Although we will not go into detail (see for example PETERSEN 1983), ergodicity and equidistribution ideas answer question 1.12 affirmatively by showing that

Under an irrational rotation ρ_α , arclength λ is the unique invariant Borel probability measure.

This gives uniqueness of Poncellet-measure μ_h when the rotation number α , coming from ellipses C and E, is irrational. But even in the case when α is rational, μ_h must still simply be the image, under φ , of the circle's arclength measure: Dilating the inner ellipse E slightly causes the rotation number to vary continuously, and only requires a small change in the affine map \mathcal{A} . Hence the Poncellet-measure for a CE-pair with rational α is the limit of Poncellet-measures from nearby irrational rotation numbers, and all these measures are the φ -image of arclength.

The Kronecker–Weyl theorem. The preceding proposition derived ergodicity of an irrational rotation directly from the density of every orbit. Exactly the same proof would show that a toral rotation is ergodic, once orbits are known to be dense. Also analogous to the 1-dimensional case of the circle, ergodicity together with Birkhoff's theorem can be used to demonstrate equidistribution, which in the multi-dimensional case is Kronecker–Weyl. Thus the fundamental idea in proving K – W is to show that rational independence of the rotation numbers implies that the orbit of the origin is dense.

Let us conclude by proving density in the 2-dimensional case, so as to illustrate how density can be lifted from the 1-dimensional case.

Warmup for the K – W theorem. Suppose numbers $\alpha, \beta, 1$ are rationally independent. Then under toral-rotation $\rho_\alpha \times \rho_\beta$, the orbit of $(0, 0)$ is dense.

Proof: Fixing ε we show that the (α, β) -orbit of $(0, 0)$ is ε -dense. Pick $N \geq 1$ such that

$$(\hat{\alpha}, \hat{\beta}) := [\rho_\alpha \times \rho_\beta]^N(0, 0)$$

is within ε of $(0, 0)$ in the torus; this, by the same Pigeon-hole argument used in (1.11). This $\hat{\alpha}$ is of the form $N\alpha$ minus an integer, and likewise for $\hat{\beta}$. Rational independence implies that $(\hat{\alpha}, \hat{\beta})$ is not $(0, 0)$ and thus

the $(\hat{\alpha}, \hat{\beta})$ -orbit of $(0, 0)$ is ε -dense in the line

$$L := \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x\hat{\beta} = y\hat{\alpha}\}$$

in the “unwrapped” torus ie., in the plane. Wrapping the plane back up, this line L winds densely around the torus, since its slope $\hat{\alpha}/\hat{\beta}$ is irrational; this again follows from rational independence. In consequence, the $(\hat{\alpha}, \hat{\beta})$ -orbit of $(0, 0)$ is ε -dense in the torus, and therefore so is the (α, β) -orbit. \square

Final remark. We know the frequency of leading-digit “9” in the doubling sequence 2, 4, 8, 16, 32, But what is the frequency of leading-digit “9” in the sequence

$$2, 4, 16, 256, 65536, \dots \quad (*)$$

where —instead of doubling— the next term is always the *square* of the current term?

Map ψ of diagram 3.2 carries squaring to the map $S(x) := 2x \pmod{1}$ on $[0, 1)$. This 2-to-1 map, an endomorphism of the circle group, preserves Lebesgue measure and is ergodic, and so Birkhoff's theorem tells us that the orbit of *almost* every point x visits each I_d interval with Lebesgue frequency. But sequence (*) asks this question of a *specific* point, $x = \log(2)$. Is $\log(2)$ a “bad” point (relative to the I_d intervals) for Birkhoff's theorem?

To this day, nobody knows . . .

REFERENCES

Library call numbers, where available, follow each reference.

- A. Avez, *Ergodic theory of Dynamical Systems* (vol. 1), University of Minnesota, Institute of Technology, 1966.
- Thøger Bang, *A solution to the “Plank Problem”*, Amer. Math. Soc. Proceedings 2 (1951). cn: *QA1.A5215*
- E. T. Bell, *Men of Mathematics*, Simon and Shuster, New York, 1937. cn: *925.1 B433m*
- H. J. M. Bos, C. Kers, F. Oort, D. W. Raven, *Poncelet's closure theorem*, Expo. Math 5 (1987), 289–364. cn: *QA 1 .E96*
- T. L. Heath, *The works of Archimedes*, Cambridge University Press, 1897. cn: *QA31A72*
- J. L. King, *Billiards inside a cusp*. Math. Intelligencer (to appear). cn: *QA1.M38*. An earlier version appears in: MSRI Preprint Series, #035-94 (1994).
- K. Petersen, *Ergodic Theory*, Cambridge University Press, 1983. cn: *QA313 .P47*
- Jean-Victor Poncelet, *Traité des propriétés projectives des figures; ouvrage utile a ceux qui s'occupent des applications de la geometrie descriptive et d'operations geometriques sûr le terrain*, Gauthier-Villars, Paris, 1866. Second edition. (First ed. published 1822) cn: *QA471.P65*
- Jean-Victor Poncelet, *Applications d'analyse et de géométrie qui ont servi, en 1822, de principal fondement au Traité des propriétés projectives des figures*, Paris, 1862.
- I. J. Schoenberg, *On Jacobi-Bertrand's proof of a Theorem of Poncelet*, in “Studies in Pure Mathematics,” Birkhauser, 1983, pp. 623–627. cn: *QA7 .S845*
- A. Tarski, *Further remarks about the degree of equivalence of polygons* (in Polish), Odbitka Z. Parametru. 2 (1932), 310–314.

Department of Mathematics
University of Florida
Gainesville, FL 32611-2082
squash@math.ufl.edu

The n -Queens Problem

Igor Rivin, Ilan Vardi, and Paul Zimmermann

1. INTRODUCTION. The n -queens problem asks how many ways can one put n queens on an $n \times n$ chessboard so that no two queens attack each other. In other words, how many points can be placed on an $n \times n$ grid so that no two are on the same row, column, or diagonal (see Figure 1).

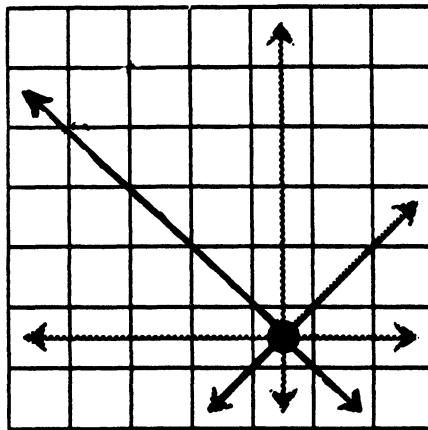


Figure 1. Chess Queen.

This question was first posed for the ordinary 8×8 chessboard as an anonymous problem [3], later attributed to Max Bezzel [1, p. 211]. The problem received wide attention, however, when posed by Franz Nauck in 1850 [22]. Writing about the problem in a letter to the astronomer Schumacher, Gauss conjectured that there were 72 solutions [9]. Soon after this 92 solutions were published which convinced Gauss that he had been incorrect. The 92 solutions are commonly represented by 12 “fundamental” solutions, that is, solutions that are not reflections and rotations of each other (see Figure 2). That 92 was the right answer, however, was not proved formally until 1874 by Dr. Glaisher [10], [27], using an idea of Günther (see Section 5). For a history of early results see [1] and the bibliography of [28].

These days the 8-queens problem is most often encountered as an exercise in introductory artificial intelligence programming courses. In fact, the n -queens problem is one of the benchmarks by which backtracking algorithms have been compared [32], [12], [11].

For the general $n \times n$ case denote by $Q(n)$ the number of solutions. It is not immediately obvious whether $Q(n) > 0$ for general n and there have been a number of independent proofs showing that $Q(2) = Q(3) = 0$, $Q(n) > 0$, $n > 3$.

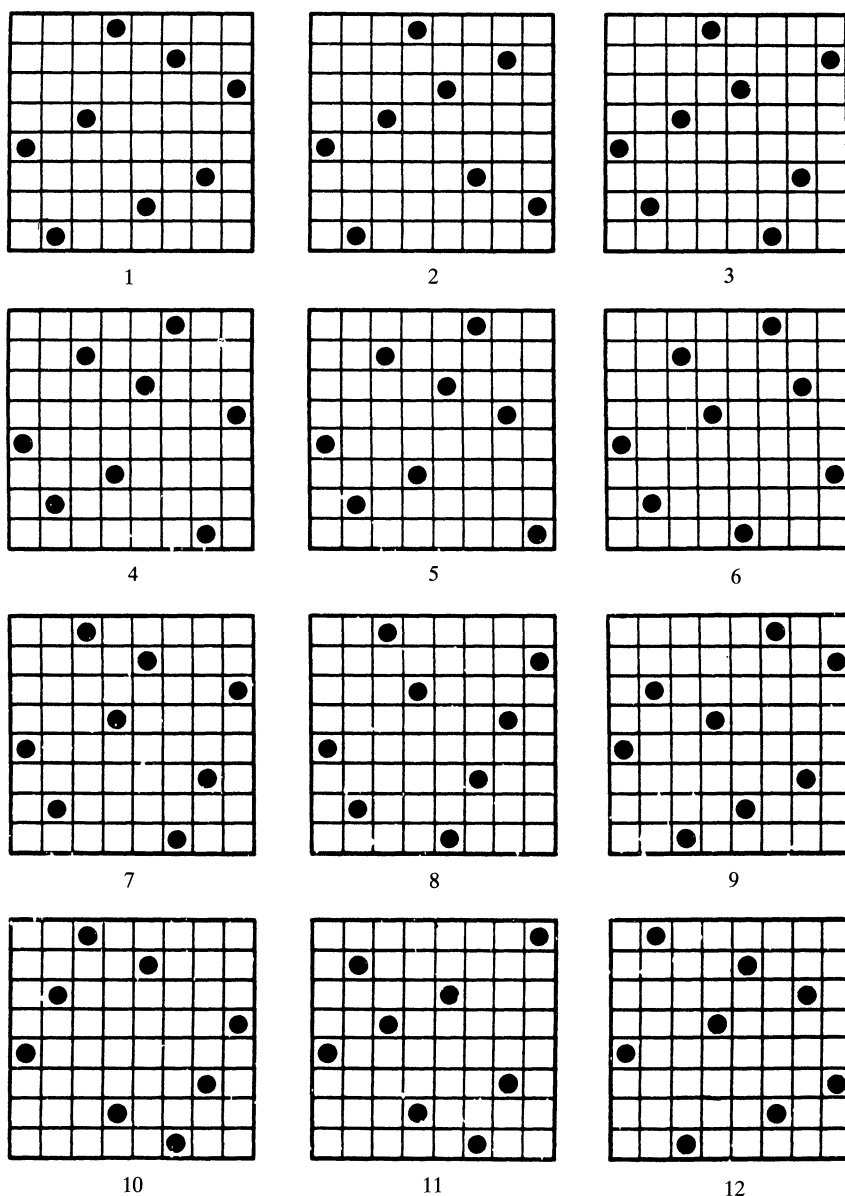


Figure 2. Fundamental solutions to 8×8 problem.

The first proof seems to be by Ahrens [1], but proofs by authors unaware of this reference appear in [34], [14], [6]. Other proofs can be found in [28], [5], [23]. It is interesting that none of these notes that $Q(1) = 1$ [35].

The precise nature of $Q(n)$ seems very difficult to understand and a more tractable problem appears to be the *toroidal n -queens problem*: How many ways can one place n -queens on an $n \times n$ chessboard so that no two queens can be on the same row, column, or extended diagonal (see Figure 3). This problem was first studied by Pólya [1, p. 363–374] who showed that $T(n) > 0$ if and only if $(n, 6) = 1$, where $T(n)$ denotes the number of $n \times n$ toroidal queens solutions.

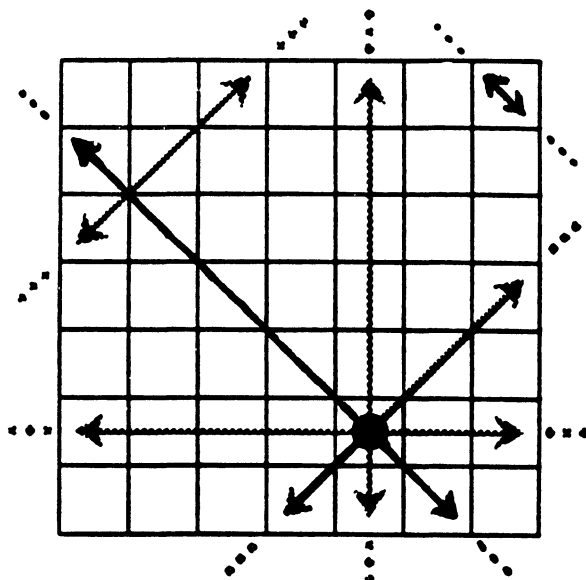


Figure 3. Toroidal Queen.

Finding closed expressions for $Q(n)$ and $T(n)$ seems to be an intractable problem, so our results deal instead with estimating the asymptotic order of these quantities.

First, note that every queen solution is also a rook solution, i.e., no two queens can be on the same column or row, and that each rook solution corresponds to a distinct permutation of $\{1, \dots, n\}$ and there are $n!$ such permutations. It follows that the trivial upper bounds are $T(n) \leq Q(n) \leq n! < e^{n \log n}$.

It appears that the only previous non-trivial lower bound is the one of Lucas [17] stating that $T(p) \geq p(p-3)$ if p is prime. In this paper we will show

Theorem 1.

- (a) Let p be a prime such that $(p-1)/2$ is not prime, then $T(p) > 2^{(p-1)/(2d)}$, where d be the smallest nontrivial divisor of $(p-1)/2$. In particular, if $p \equiv 1 \pmod{4}$, then $T(p) > 2^{(p-1)/4}$, but in general one only has the bound $T(p) > 2^{\sqrt{(p-1)/2}}$.
- (b) If n is divisible by a prime $\equiv 1 \pmod{4}$ then $T(n) > 2^{n/5}$.

Note that the above results also hold for $Q(n)$ since $Q(n) \geq T(n)$, while (b) holds for almost all n , since the set of integers not divisible by a prime $\equiv 1 \pmod{4}$ has density zero [13].

We have not been able to find super exponential bounds for $Q(n)$ and $T(n)$ but we believe

Conjecture 1.

$$\lim_{\substack{n \rightarrow \infty \\ (n, 6) = 1}} \frac{\log T(n)}{n \log n} = \alpha > 0, \quad \lim_{n \rightarrow \infty} \frac{\log Q(n)}{n \log n} = \beta > 0.$$

Finally, $T(n)$ can also be thought of as the number of arrays of non-negative integers with row, column, and broken diagonals summing to one, i.e., a *pandiagonal magic square* [27] with common sum equal to one. Following ideas of [30] we propose

Conjecture 2. *The generating function $\sum_{n=1}^{\infty} (T(n)/n!)x^n$ has a closed form.*

2. SURVEY OF PREVIOUS RESULTS. First, write a queens solution as a function $f(k)$, $k = 0, \dots, n - 1$, so that the k 'th queen is placed at the $(k, f(k))$ coordinate of the chessboard. It follows immediately that f represents a toroidal solution if and only if $k \mapsto f(k)$, $k \mapsto f(k) + k \pmod{n}$, and $k \mapsto f(k) - k \pmod{n}$ are all one to one.

Similarly, $f(k)$ represents a (not necessarily toroidal) queens solution if and only if $k \mapsto f(k)$, $k \mapsto f(k) + k$, and $k \mapsto f(k) - k$ are one to one.

We now present an elegant proof [5], [2], that there is always a queens solution for $n > 3$. The proof splits up according to the residue class of $n \pmod{6}$.

- (a) If $n = 6m + 1$ or $n = 6m + 5$ then $(n, 6) = 1$ and one lets $f(k)$ be given by $f(k) = 2k \pmod{n}$. This is clearly a toroidal solution (thus an ordinary solution). Note that this is what one would ordinarily consider as putting queens one "knight's move" apart.
- (b) If $n = 6m$ or $n = 6m + 4$ then one takes the solution of (a) for the $(n + 1) \times (n + 1)$ board and removes the queen in the $(0, 0)$ position (i.e., the leftmost column and bottom row). The resulting position is an n solution.
- (c) If $n = 6m + 2$ or $n = 6m + 3$ first construct a $6m + 2$ solution as follows: Put a queen at $(k, f(k))$, where

$$f(k) = \begin{cases} 2k + (n - 2)/2 \pmod{n}, & \text{if } 0 \leq k \leq (n - 2)/2 \\ n - 1 - f(n - 1 - k), & \text{if } n/2 \leq k \leq n - 1 \end{cases}.$$

One checks easily that this is a solution. Note that this is a straightforward generalization of solution (10) in Figure 2 for the 8×8 case.

Since this solution does not have a queen on the main diagonal, one can construct a $6m + 3$ solution by adding a row and column to the edge of the board and putting a queen on the new corner. \square

Turning to $T(n)$, we prove Pólya's result characterizing n for which $T(n) > 0$. This is also proved in [5], [19], [2]. An extension is given in [20].

Let $(n, 6) = 1$, then, as before, $f(k) = 2k \pmod{n}$ is a toroidal solution, so $T(n) > 0$. Conversely, one shows that $T(n) > 0$ implies that n is not divisible by 2 or 3.

Assume that $f(k)$ represents an $n \times n$ toroidal solution so $f(k) - k \pmod{n}$ is a permutation of $0, \dots, n - 1$ and

$$\sum_{k=0}^{n-1} (f(k) - k) \equiv \sum_{k=0}^{n-1} k = \frac{n(n-1)}{2} \pmod{n}.$$

But this sum is also

$$\sum_{k=0}^{n-1} (f(k) - k) = \sum_{k=0}^{n-1} f(k) - \sum_{k=0}^{n-1} k = 0,$$

since $f(k)$ is a permutation of $0, \dots, n - 1$. Therefore n divides $n(n - 1)/2$ and it follows that n is odd.

One similarly shows that n is not divisible by 3 by using the more elaborate sum

$$\sum_{k=0}^{n-1} (f(k) - k)^2 + \sum_{k=0}^{n-1} (f(k) + k)^2 - 4 \sum_{k=0}^{n-1} k^2 \equiv 0 \pmod{n}. \quad \square$$

The next result, also due to Pólya [1], shows how to *compose* solution, i.e., if an $m \times m$ solution and an $n \times n$ solution are given then one tries to construct an $mn \times mn$ solution by placing a copy of the $m \times m$ solution where each queen appears in the $n \times n$ solution (see Figure 4).

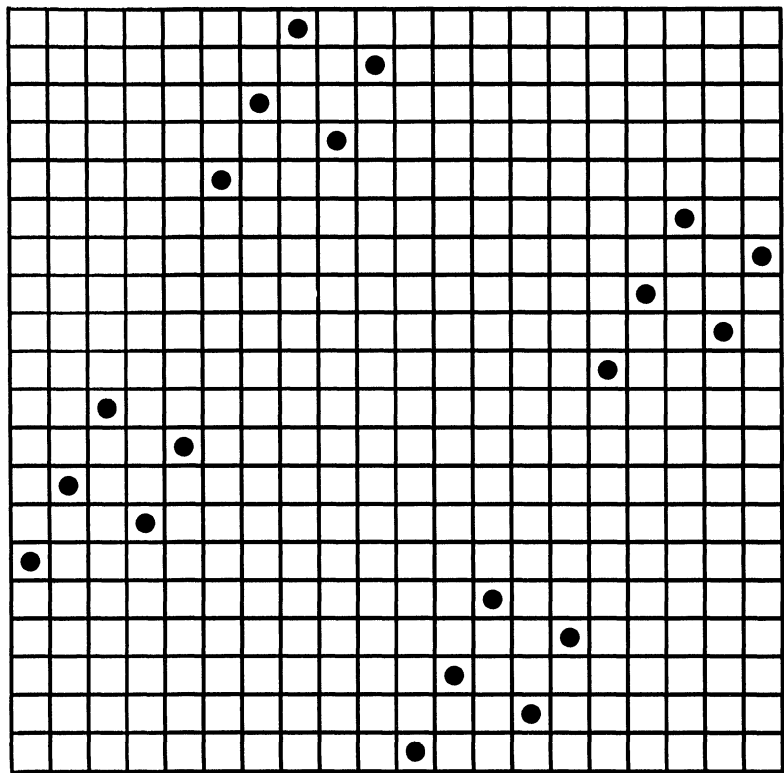


Figure 4. Composed solution.

Let $m, n > 3$, where $(n, 6) = 1$. Then if g is a toroidal $n \times n$ solution and f is an ordinary $m \times m$ solution, then one can compose these to an $mn \times mn$ solution.

Proof: Every integer mod mn can be written uniquely as $an + b$, $a = 0, \dots, m - 1$, $b = 0, \dots, n - 1$. The claim is that $h(an + b) = f(a)n + g(b)$ is an $mn \times mn$ solution. For example, checking the condition that $h(k) + k$ is one to one: The assumption

$$(1) \quad h(an + b) + an + b = h(a'n + b') + a'n + b'$$

is seen to imply $g(b) + b \equiv g(b') + b' \pmod{n}$, which gives $b = b'$ since g is a $n \times n$ toroidal solution. Equation (1) then gives that

$$f(a)n + an = f(a')n + a'n,$$

so $a = a'$ since f is a queens solution. The other cases are exactly similar, giving the result.

We now turn to the Lucas estimate $T(p) \geq p(p-3)$ which is proved by noting that for p a prime, each pair $a \neq 0, \pm 1 \pmod{p}$ and $b = 0, \dots, p-1$, generates the distinct toroidal solution $f(k) = ak + b$, and the number of such solutions is $p(p-3)$.

We end this section by mentioning results that are tangentially related to the n -queens problem.

(i) Upper bounds for $T(n)$ might be obtained by replacing the toroidal queen with a *toroidal semiqueen*, a piece that moves like a toroidal queen but cannot travel on negative diagonals. The toroidal semiqueen problem can be expressed very simply in terms of permanents and this question was studied by I. Rivin and I. Vardi in [33, Chapter 6].

(ii) A simpler question than the n -queens problem is to compute how many ways one can place k non attacking queens on an $n \times n$ chessboard. Call this number $Q_k(n)$. For $k = 2, 3$ there are closed forms [1], [28],

$$Q_2(n) = \frac{n(n-1)(n-2)(3n-1)}{6},$$

$$Q_3(n) = \frac{(n-1)(n-3)(2n^4 - 12n^3 + 25n^2 - 14n + 1)}{12}.$$

In general, one can show [31, Problem 4.15] that for fixed k the generating function

$$\sum_{n=1}^{\infty} Q_k(n) x^n$$

is a rational function.

(iii) The ideas in Pólya's paper have been used to give another proof of Fermat's result that every prime $\equiv 1 \pmod{4}$ is a sum of two squares [16].

3. COMPOSING NEW SOLUTIONS. Examining a typical composition constructed by Theorem 3 one notes that there are large areas of the board that have no queens (see Figure 4).

We noted that these regions can be used effectively to construct many more solutions with only a slight variant of the basic composition idea.

Theorem 2. *Let $m, n > 3$ be given, $(n, 6) = 1$, and let $f_1, f_2, \dots, f_{Q(m)}$ be all $m \times m$ queens solutions, and let g be a toroidal $n \times n$ solution. Then for each map $\pi: \{0, \dots, n-1\} \rightarrow \{1, \dots, Q(m)\}$ the function $h(an+b) = f_{\pi(b)}(a)n + g(b)$ gives a distinct $mn \times mn$ solution.*

Proof: The proof that each of these gives a queens solution is exactly as in the previous section, while the fact that each solution is distinct is clear from the definition. \square

Corollary 1. *Let $(n, 6) = 1, m \geq 3$ then $Q(mn) > [Q(m)]^n T(n)$. In particular, if N is a number divisible by 5, and $(N, 6) = 1$, then $Q(N) > 4^{N/5}$.*

Proof: The first part follows directly from counting the number of solutions generated by Theorem 2. The second result follows by letting $n = 5, m = N/5$, and noting that $Q(5) = 4$. One then checks the special cases $N = 10, 15$. \square

Remark. Corollary 1 gives the first example of a set of n 's for which $Q(n)$ grows faster than a polynomial in n .

4. USING THE MULTIPLICATIVE STRUCTURE OF Z/pZ . In the previous section we constructed $mn \times mn$ solutions out of $m \times m$ solutions and $n \times n$ solutions. Such a method will not work for the case of $p \times p$ chessboards, when p is prime, so constructing an exponential number of solutions in this case requires a new idea.

The basic idea of composition was to generate solutions using an additive subgroup of Z/nZ . We will now use the multiplicative structure by constructing "quasi-linear" solutions of the form $f(k) = c_k k$, where c_k is constant on cosets of a multiplicative subgroup of $(Z/pZ)^*$.

The simplest case is when $p \equiv 1 \pmod{4}$. It is well known [13, page 85] that for such p there is a number $i \pmod{p}$ with the property that $i^2 \equiv -1 \pmod{p}$.

Consider an equivalence relation on $\{1, \dots, p-1\}$ by $a \sim b$ if $a = i^k b$, for some k . This defines $(p-1)/4$ equivalence classes $\langle a_1 \rangle, \langle a_2 \rangle, \dots, \langle a_{(p-1)/4} \rangle$ for some sequence numbers $0 < a_1, \dots, a_{(p-1)/4} < p$.

Now consider $f(k) = c_k k$, where $c_k = i$ or $c_k = 1/i$, and c_k is constant on each set $\langle k \rangle$. The claim is that $f(k)$ is a toroidal queens solution.

Theorem 3 (i). *Each map $\sigma: \{1, 2, \dots, (p-1)/4\} \rightarrow \{\pm 1\}$ yields a distinct toroidal $p \times p$ solution.*

Proof: For each σ identify $\{1, \dots, (p-1)/4\}$ with the distinct classes $\langle a_1 \rangle, \dots, \langle a_{(p-1)/4} \rangle$ and define f by $f(k) = i^{\sigma(\langle k \rangle)} k$.

To see that f is a toroidal solution we check the three conditions of Section 2:

(a) Assume that $f(k) = f(k')$, then

$$i^{\sigma(\langle k \rangle)} k = i^{\sigma(\langle k' \rangle)} k' \Rightarrow k = i^{\pm 1} k' \Rightarrow \langle k \rangle = \langle k' \rangle \Rightarrow k = k'$$

so f is one to one.

(b) Assume that $f(k) + k = f(k') + k'$ then $i^{\sigma(\langle k \rangle)} k + k = i^{\sigma(\langle k' \rangle)} k' + k'$ so

$$k(1 + i^{\pm 1}) = k'(1 + i^{\mp 1})$$

for one of four choices of $\pm 1, \mp 1$. All these cases lead to $\langle k \rangle = \langle k' \rangle$, for example, if $k(1 + 1/i) = k'(1 + i)$ then the identity $1 + 1/i = (1 + i)/i$ gives $k = ik'$. It follows that $\langle k \rangle = \langle k' \rangle$ so $k = k'$.

(c) If $f(k) - k = f(k') - k'$, then one gets that $k = k'$ as in part (b). The only difference is that $i^2 \equiv -1 \pmod{p}$ is needed to take case of the case when $k(i - 1) = k'(1/i - 1)$ and $k(1/i - 1) = k'(i - 1)$.

Finally, it is routine to check that for each distinct σ one gets a distinct f . \square

In the general case let q be the smallest divisor of $p-1$ that is even and greater than two. It is known [13] that $x^q = 1$ has q solutions mod p and that these form a cyclic group. Let ξ be a generator and, as before, define an equivalence relation by $a \sim b$ if $a = \xi^k b$ for some k . This gives $(p-1)/q$ equivalence classes $\langle a_1 \rangle, \dots, \langle a_{(p-1)/q} \rangle$, where $\langle a \rangle$ represents the equivalence class of a (i.e., $\{a, \xi a, \xi^2 a, \dots, \xi^{q-1} a\}$). The result is

Theorem 3 (ii). *Each map $\sigma: \{1, \dots, (p-1)/q\} \rightarrow \{\pm 1\}$ leads to a distinct toroidal $p \times p$ solution.*

Proof: The proof proceeds exactly as in part (i) of the theorem. It is important to note that as in part (c) above, one needs to have $\xi^j = -1$ for some j . This explains why q must be chosen to be an *even* divisor of $p - 1$. □

Theorem 1(a) now follows immediately by counting the number of solutions generated by Theorem 3.

Remark 1. We have found no improvement on the lower bound $p(p - 3)$ for primes of the form $p = 2q + 1$, q prime (Cunningham primes).

Remark 2. For a given q dividing $p - 1$, the solutions constructed by this method have the same cycle structure when taken as permutations, i.e., a product of $(p - 1)/q$ cycles of length q .

Remark 3. This method of constructing solutions can be used to give more complicated forms of compositions of solutions. For simplicity consider two distinct primes $p_1, p_2 \equiv 1(\text{mod } 4)$ (the general case is similar).

Theorem 4. *For each map $\sigma: \{1, \dots, (p_1 - 1)/4\} \rightarrow \{1, \dots, T(p_2)\}$ and $\pi: \{0, \dots, p_2 - 1\} \rightarrow \{1, \dots, 2^{(p_1-1)/4}\}$ there is a distinct $p_1 p_2 \times p_1 p_2$ toroidal solution.*

Proof: Let $f_1, \dots, f_{T(p_2)}$ be the toroidal $p_2 \times p_2$ solutions and $g_1, \dots, g_{2^{(p_1-1)/4}}$ be the solutions as constructed in Theorem 6 (i). One can write each number $(\text{mod } p_1 p_2)$ uniquely as $ap_1 + bp_2$ where $a = 0, \dots, p_2 - 1, b = 0, \dots, p_1 - 1$. It can then be shown that for each σ, π the function

$$h(ap_1 + bp_2) = f_{\sigma^{(\langle b \rangle)}}(a)p_1 + g_{\pi(a)}(b)p_2.$$

gives a distinct toroidal solution. □

Counting the number of solutions generated by Theorem 4, one gets that

$$T(p_1 p_2) \geq T(p_2)^{(p_1-1)/4} 2^{p_2(p_1-1)/4}.$$

This can be extended to more complicated compositions for products of more than two primes. A computation shows that in the limit this gives the lower bound $T(n) > 2^{(1-\varepsilon)n/3}$, where $\varepsilon \rightarrow 0$ as the number of prime factors of n that are $\equiv 1 \pmod{4}$ goes to infinity.

We now turn to the proof of Theorem 1 (b). Consider $n, (n, 6) = 1$, and n is divisible by a prime $p \equiv 1 \pmod{4}$. It follows from Corollary 1 and $p \geq 5$ that

$$T(n) = T((n/p)p) > T(p)^{n/p} T(n/p) > 2^{n(p-1)/(4p)} \geq 2^{n/5}. \quad \square$$

Remark. The reader may have noted that our techniques have been elementary. It is an interesting question why nontrivial lower bounds have escaped the large literature on this subject. We believe that there are two reasons for this.

The first comes for the original formulation of the problem on the 8×8 chessboard. Note that in the proof of Theorem 1 the hardest case was for numbers $\equiv 2 \pmod{6}$, since they cannot be reached from the more tractable toroidal problem. This might also explain why the toroidal problem has not been extensively investigated.

The second reason is the emphasis on classifying “fundamental” solutions, i.e., solutions that are not rotations or reflections of each other. This problem is difficult even in the much simpler case of the rook’s problem, and counting the number of fundamental solutions has proved to be nontrivial [18], [4], [21], [26]. Note that fundamental solutions under toroidal symmetries for the toroidal case are very easy to classify—they are the ones with a queen at (0, 0).

5. COMPUTATIONAL RESULTS. As mentioned in the introduction there has been much interest in the computation of $Q(n)$ and $T(n)$ using backtracking algorithms. A different algorithm has been advanced by Igor Rivin and Ramin Zabih [24]. Their idea is similar to a method proposed by Günther: Consider independent variables $X_0, \dots, X_{2n-2}, Y_{-n+1}, \dots, Y_{n-1}$ and the matrix $\|X_{i+j}Y_{i-j}\|$, then the squarefree term in X, Y of the *permanent* of this matrix (determinant with no minus signs) gives the number of queens solutions. As before the toroidal case is much cleaner. One has $2n$ variables and considers the squarefree term of the permanent of $\|X_{i+j \bmod n}Y_{i-j \bmod n}\|$.

To estimate the running time of this method in the toroidal case, note that the standard methods for evaluating a permanent give a running time of about 2^n multiplications, and the terms are squarefree polynomials in $2n$ variables and of degree $\leq 2n$. It follows that there are at most 2^{2n} terms. Since a multiplication takes time about 8^n , this gives a running time on the order of 16^n , but with space requirement of 4^n (the running time can be reduced to 8^n [25]).

Since backtracking algorithms generate all solutions these take at least $T(n)$ steps to compute $T(n)$. It follows that a lower bound $T(n) > \gamma^n$, where $\gamma > 8$, would show that the Rivin-Zabih algorithm always runs faster than backtracking (if Conjecture 1 holds, then this algorithm will be much faster than backtracking).

On the other hand, backtracking takes very little space so it is still the more practical method for computing large values of $T(n)$ and $Q(n)$, and all the values in Figure 5 were computed this way ($Q(19)$ and $Q(20)$ were computed by A. Shapira [29]). For example, the value $T(23) = 128850048$ was computed using backtracking and a number of implementation shortcuts. Toroidal symmetries

n	$T(n)$	$\log T(n)/(n \log n)$	$Q(n)$	$\log Q(n)/(n \log n)$
4			2	0.125
5	10	0.286	10	0.286
6			4	0.129
7	28	0.245	40	0.271
8			92	0.272
9			352	0.297
10			724	0.286
11	88	0.170	2680	0.299
12			14200	0.321
13	4524	0.252	73712	0.336
14			365596	0.347
15			2279184	0.360
16			14772512	0.372
17	140692	0.246	95815104	0.382
18			666090624	0.391
19	820496	0.243	4968057848	0.399
20			39029188884	0.407
23	128850048	0.259		

Figure 5. Values of $T(n)$, $Q(n)$.

reduced the number to be computed to 1482252 solutions. A further saving was to eliminate impossible consecutive triplets. The computation was done in LeLisp as a distributed computation over a network of 20 Suns at INRIA, Rocquencourt, and took 267 days of CPU time.

Note that the computational evidence supports Conjecture 1 since the values of $\log Q(n)/(n \log n)$ and $\log T(n)/(n \log n)$ seem to be monotonically increasing.

REFERENCES

1. W. Ahrens, *Mathematische Unterhaltungen und Spiele, Vol. 1*, B. G. Teubner, Leipzig 1921.
2. J. D. Beasley, *The Mathematics of Games*, Oxford University Press, Oxford 1990.
3. Berliner Schachgesellschaft, **3** (1848), p. 363.
4. S. Chowla, I. N. Herstein, and K. Moore, *On recursions connected with symmetric groups I*, Canadian J. of Math. **3** (1951), 328–334.
5. D. Clark, *A combinatorial theorem on circulant matrices*, This Monthly **92** (1985), 725–729.
6. B.-J. Falkowski and L. Schmitz, *A note on the queen's problem*, I.P.L. (= Information Processing Letters) **23** (1986), 39–46.
7. L. R. Foulds and D. G. Johnston, *An application of Graph Theory and Integer Programming: Chessboard non-attacking puzzles*, Math. Gazette **57** (1984), 95–104.
8. M. Gardner, *The unexpected Hanging and other Mathematical Diversions*, Simon and Schuster, New York 1986.
9. C. F. Gauss, Letters to H. C. Schumacher, September 12–27, 1850, *Werke, Vol. 12*, Springer, Berlin 1929, p. 20–28.
10. Dr. J. W. L. Glaisher, Philosophical Magazine, **18** (1874), 457–467.
11. S. Golomb and L. Baumert, *Backtrack programming*, J.A.C.M. **12** (1965), 516–524.
12. R. Haralick and G. Elliott, *Increasing Tree Search Efficiency for Constraint Satisfaction Problems*, Artificial Intelligence **14** (1980), 263–313.
13. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford 1988.
14. E. J. Hoffman, J. C. Loessi, and R. C. Moore, *Construction for the solution of the n-queens problem*, Math. Mag. **42** (1969), 66–72.
15. M. Kraitchik, *Mathematical Recreations*, Dover, New York 1953.
16. L. Larson, *A theorem about primes proved on a Chessboard*, Math. Mag. **50** (1977), 69–74.
17. E. Lucas, *Récréations Mathématiques, Vol. I*, Albert Blanchard, Paris 1977.
18. E. Lucas, *Théorie des Nombres*, Albert Blanchard, Paris 1961.
19. P. Monsky, Problem E 2698, This Monthly **85** (1978), p. 116.
20. P. Monsky, Problem E 3162, This Monthly **93** (1986), p. 566.
21. L. Moser and M. Wyman, *On the solution of $x^d = 1$ in symmetric groups*, Canadian J. of Math. **7** (1955), 159–168.
22. F. Nauck, *Illustrierten Zeitung*, **14**, p. 352, June 1, 1850; **15**, p. 182, September 21, 1850; **15**, p. 207, September 28, 1850.
23. M. Reichling, *A Simplified Solution of the N Queens' Problem*, I.P.L. **25** (1987), 253–255.
24. I. Rivin and R. Zabih, *An Algebraic Approach to Constraint Satisfaction Problems*, Proceedings of the Int. Joint Conf. on A.I., Detroit 1989.
25. I. Rivin and R. Zabih, *A dynamic programming solution to the N-queens problem*, I.P.L. **41** (1992), 253–256.
26. R. W. Robinson, *Counting the arrangements of Bishops*, in “Combinatorial Mathematics IV, Adelaide 1975,” L. Notes in Math. **560** (1976), Springer-Verlag 1976, 198–214.
27. W. W. Rouse Ball and H. S. M. Coxeter, *Mathematical Recreations and Essays*, Dover 1987.
28. A. Sainte-Laguë, *Les Réseaux (ou Graphes)*, Mém. Des Sc. Math. **18**, Gauthier-Villars, Paris 1926.
29. A. Shapira, Personal communication, August 1992.
30. R. P. Stanley, *Combinatorics and Commutative Algebra*, Birkhäuser 1983.
31. R. P. Stanley, *Enumerative Combinatorics, Vol. 1*, Wadsworth 1986.
32. H. S. Stone and J. M. Stone, *Efficient Search Techniques-An Empirical study of the N-Queens Problem*, IBM Technical Report, T. J. Watson Research Center.
33. I. Vardi, *Computational Recreations in Mathematica*, Addison Wesley 1990.

34. A. M. Yaglom and I. M. Yaglom, *Challenging Mathematical Problems with Elementary Solutions*, Vol. I, Dover 1987.
35. R. Zabih, Personal communication 1988.

Rivin:

*Mathematics Department
The University of Melbourne
Parkville, Victoria 3052
rivin@geom.umn.edu*

Vardi:

*MSRI
1000 Centennial Drive
Berkeley, CA 94720
ilan@leland.stanford.edu*

Zimmermann:

*INRIA Lorraine
Nancy, France
paul.zimmerman@inria.fr*

The Prince of Algebra

Madam Professor,
Let me introduce myself—
I'm Albert James,
whom you may know
by my test score
that's lower than my age.

Your algebra tests
are too long for me
in fifty minutes,
but I am proud
of my attendance—
I never miss class,
never come late.

I am preparing
for a new career.
For thirty years I was
with the Postal Service
never absent,
never late.

Your mathematics
is important!
It runs the clock
which runs the mail.
Now I train to be
a first grade teacher.

I will teach
mathematics
by punctuality
and perfect attendance.

*From Intersections: Poems by JoAnne Growney,
Kadet Press, Bloomsburg, PA, 1993, p. 52–53.*

What's the Difference Between Cantor Sets?

Roger L. Kraft

In this article, we will look at a family of sets called the middle- α Cantor sets and we will try to show that not all of these sets are of the same “size.” We will do this without computing any of the myriad kinds of fractal dimensions. Instead, we will prove three theorems about what are called the difference sets of the middle- α Cantor sets and these theorems will provide the evidence we need to conclude that some middle- α Cantor sets are “larger” than others. The proofs of these theorems will use methods that emphasize the ideas of self-similarity, rescaling and renormalization.

The middle- α Cantor sets are a straightforward generalization of the classical middle third Cantor set (see [B, pp. 33–36] or [E, pp. 1–6]). To define a middle- α Cantor set in the interval $[0, 1]$, first choose $\alpha \in (0, 1)$, let $\beta = (1 - \alpha)/2$ and then define the affine maps

$$T_0(x) = \beta x \quad \text{and} \quad T_1(x) = \beta x + (1 - \beta).$$

Let $I_0 = [0, 1]$ and then, for $n \geq 1$, inductively define

$$I_n = T_0(I_{n-1}) \cup T_1(I_{n-1}).$$

Then the middle- α Cantor set in the interval $[0, 1]$ is defined as

$$\Gamma_\alpha = \bigcap_{n=0}^{\infty} I_n.$$

Let's look at this definition more carefully. Notice that $T_0(I_0) = [0, \beta]$ and $T_1(I_0) = [1 - \beta, 1]$, so $I_1 = [0, \beta] \cup [1 - \beta, 1]$. The “hole” in I_1 has length $1 - 2\beta = \alpha$. So another way of describing I_1 is to say that we remove from the middle of I_0 the open interval of length α , leaving two closed intervals of length β . Now notice that $T_0(I_1) = [0, \beta^2] \cup [\beta(1 - \beta), \beta]$ and that $T_1(I_1) = [1 - \beta, \beta^2 + (1 - \beta)] \cup [(1 - \beta) + \beta(1 - \beta), 1]$. So we get I_2 from I_1 by removing from the middle of each component of I_1 an open interval of length $\alpha\beta$. Each of the four components of I_2 has length β^2 . In general, I_n is a disjoint union of 2^n closed intervals of length β^n and we get I_{n+1} by removing from the middle of each component of I_n an open interval of length $\alpha\beta^n$. The collection $\{I_n\}_{n=0}^{\infty}$ is a nested sequence of compact subsets of $[0, 1]$ and it has the finite intersection property, so by the compactness of $[0, 1]$, Γ_α is a nonempty set. Each Γ_α is a compact, nowhere dense, perfect subset of the real line and hence is a Cantor set. (How would you modify T_0 and T_1 to define a middle- α Cantor set in the interval $[a, b]$?)

Fix a choice of $\alpha \in (0, 1)$ and let x be a point in Γ_α . Define the *address* in Γ_α of x to be a sequence $\{s_0, s_1, s_2, \dots\}$ where each s_i is either 0 or 1. To determine s_i , consider x as a point in I_{i+1} and if x is to the left of the nearest hole of length

$\alpha\beta^i$, then $s_i = 0$ and if x is to the right of the nearest hole of length $\alpha\beta^i$, then $s_i = 1$. The primary usefulness of the address of a point x from Γ_α is the fact, which you should check, that x has a unique representation of the form

$$x = \sum_{i=0}^{\infty} s_i \beta^i (1 - \beta) \quad \text{where } s_i \in \{0, 1\},$$

and the “digits” in this representation are the terms of the address. (Hint: The 2^n points of the form $\sum_{i=0}^{n-1} s_i \beta^i (1 - \beta)$ are the left-hand endpoints of the 2^n components of I_n .) You should also check that when $\alpha = \beta = 1/3$, this representation is equivalent to the standard ternary representation of the points in the middle third Cantor set.

You may have noticed that the number β is more useful than α in describing Γ_α . This is because β^{-1} is a scaling factor for Γ_α . By this we mean that $\beta^{-1}(\Gamma_\alpha \cap [0, \beta]) = \Gamma_\alpha$ (where, if γ is a real number and A is a subset of the real line, then $\gamma A = \{\gamma x | x \in A\}$). In the rest of this article, when referring to a middle- α Cantor set, we will freely go back and forth between the two numbers $\beta = (1 - \alpha)/2$ and $\alpha = 1 - 2\beta$.

• Compare the middle-9/10 and the middle-1/10 Cantor sets. One seems “larger” than the other. But how can we demonstrate this? We could try comparing the number of points in the two sets but it’s not too hard to show that all the middle- α Cantor sets have the same cardinality (in fact, any two Cantor sets are homeomorphic; see [HY, pp. 97–100]). So we can’t use cardinality to compare the sizes of different middle- α Cantor sets. Another way to compare the sizes of two sets is to compare the values of their Lebesgue measure. But we’ll now show that each of the middle- α Cantor sets has Lebesgue measure zero. Choose $\alpha \in (0, 1)$. Each set I_n covers Γ_α . I_n contains 2^n components, each of length β^n . So the total length of I_n is $2^n \beta^n = (2\beta)^n$. So the total length of Γ_α is less than or equal to $(2\beta)^n$ for all n . Now $\alpha \in (0, 1)$ implies that $\beta \in (0, 1/2)$, so $2\beta < 1$ and therefore the total lengths of the I_n go to zero as n goes to infinity. And this is what it means to say that the Lebesgue measure of Γ_α is zero. So we can’t use Lebesgue measure to compare the sizes of the middle- α Cantor sets. (It is not true, however, that every Cantor set has measure zero; see [B, pp. 63–64].)

So we can’t use cardinality or Lebesgue measure to distinguish between different middle- α Cantor sets. But it would be nice if we could find some way to use these two simple notions to demonstrate that the middle-1/10 Cantor set is “larger” than the middle-9/10 Cantor set. And that is exactly what we will be able to do, by applying cardinality and Lebesgue measure to the *difference set*, $\Gamma_\alpha - \Gamma_\alpha$, of a middle- α Cantor set. The name and notation for the difference set come from the definition,

$$\Gamma_\alpha - \Gamma_\alpha = \{x - y | x, y \in \Gamma_\alpha\}.$$

Here is a more “dynamical” way of defining the difference set,

$$\Gamma_\alpha - \Gamma_\alpha = \{t | \Gamma_\alpha \cap (\Gamma_\alpha + t) \neq \emptyset\}$$

where $\Gamma_\alpha + t = \{x + t | x \in \Gamma_\alpha\}$. In this second definition, $\Gamma_\alpha + t$ is a translate of Γ_α and $\Gamma_\alpha - \Gamma_\alpha$ tells us which of these translates intersect with Γ_α . Imagine a copy of Γ_α moving down the number line with constant velocity one such that it coincides with Γ_α when $t = 0$. The difference set describes those times when the moving copy of Γ_α intersects with Γ_α . Notice that for any choice of $\alpha \in (0, 1)$, $\Gamma_\alpha - \Gamma_\alpha \subset [-1, 1]$.

How can difference sets be used to show that one middle- α Cantor set is larger than another? Intuitively, the larger a set is, the more often it should intersect with a translate of itself, so the larger its difference set should be. Also, the larger a set is, the larger we would expect the intersection that set has with its translates to be. So what we will do is apply cardinality and Lebesgue measure to the difference sets $\Gamma_\alpha - \Gamma_\alpha$ and to the intersections $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ for $t \in \Gamma_\alpha - \Gamma_\alpha$. We will show that, as α decreases, the Lebesgue measure of $\Gamma_\alpha - \Gamma_\alpha$ will increase, and the minimum cardinality of $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ for $t \in \Gamma_\alpha - \Gamma_\alpha$ will increase. However, the price we pay for using such simple concepts as Lebesgue measure and cardinality is that we can only demonstrate changes in the size of Γ_α for a few values of α . We will not be able to demonstrate a continuous increase in the size of the middle- α Cantor sets as α decreases, which is really the case.

Our first theorem is about the measure of the difference sets. We will state the theorem in terms of β instead of α . Notice that the sets Γ_α “grow” as β increases.

Theorem 1. *If $\beta < 1/3$, then $\Gamma_\alpha - \Gamma_\alpha$ is a Cantor set of measure zero. If $\beta \geq 1/3$, then $\Gamma_\alpha - \Gamma_\alpha = [-1, 1]$.*

In other words, if $\beta < 1/3$, then $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is empty for almost all $t \in [-1, 1]$ and if $\beta \geq 1/3$, then $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is nonempty for all $t \in [-1, 1]$. So there is a significant change in the way that the middle- α Cantor sets act when β (or α) crosses $1/3$. We will see later that this change is even more dramatic than is indicated by this theorem.

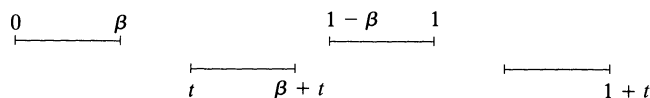
Proof: A very elegant proof by picture for the case when $\beta = 1/3$ can be found in [B, p. 110]. That proof can easily be seen to generalize to all $\beta \in [1/3, 1/2)$. So we will leave it to the reader to look up the proof of this part of the theorem.

We will prove the part of the theorem where $\beta < 1/3$. We will use the fact (which you should prove) that for $\alpha \in (0, 1)$,

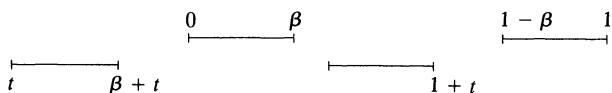
$$\Gamma_\alpha - \Gamma_\alpha = \bigcap_{n=0}^\infty (I_n - I_n).$$

We will show, by induction, that $I_n - I_n$ is a disjoint union of 3^n closed intervals each of which has length $2\beta^n$. This implies that $\Gamma_\alpha - \Gamma_\alpha$ has measure zero by an argument very similar to the one used to prove that Γ_α has measure zero. (Each $I_n - I_n$ is a cover of $\Gamma_\alpha - \Gamma_\alpha$ and the total length of $I_n - I_n$ is $3^n 2\beta^n = 2(3\beta)^n$ which goes to zero as n goes to infinity since $\beta < 1/3$.) The fact that $\Gamma_\alpha - \Gamma_\alpha$ is a Cantor set will follow from the way that $I_{n+1} - I_{n+1}$ is derived from $I_n - I_n$.

To make clearer the geometric structure of $\Gamma_\alpha - \Gamma_\alpha$, we will begin the induction with $n = 1$, rather than with $n = 0$. If $\beta < t < 1 - 2\beta$, then it's easy to see that $I_1 \cap (I_1 + t) = \emptyset$ (notice that if $\beta < 1/3$, then $\alpha > \beta$; now consider the following picture of I_1 and $I_1 + t$).



And if $-1 + 2\beta < t < -\beta$, then we again have $I_1 \cap (I_1 + t) = \emptyset$.



For any other choice of $t \in [-1, 1]$ we will have $I_1 \cap (I_1 + t) \neq \emptyset$. So $I_1 - I_1$ is the disjoint union of three closed intervals, and each of these intervals has length 2β , i.e.,

$$I_1 - I_1 = [-1, -1 + 2\beta] \cup [-\beta, \beta] \cup [1 - 2\beta, 1].$$

Now suppose that $I_n - I_n$ is the disjoint union of 3^n closed intervals, each of length $2\beta^n$.

For $n \geq 1$ let $I_n^L = I_n \cap [0, \beta]$ and let $I_n^R = I_n \cap [1 - \beta, 1]$, so I_n^L and I_n^R are the left and right “halves” of I_n . Because of the self-similarity of Γ_α , we have $I_n = \beta^{-1}I_{n+1}^L$. So

$$\beta^{-1}I_{n+1}^L - \beta^{-1}I_{n+1}^L = I_n - I_n$$

or

$$I_{n+1}^L - I_{n+1}^L = \beta(I_n - I_n)$$

and so the induction hypothesis implies that $I_{n+1}^L - I_{n+1}^L$ is the disjoint union of 3^n closed intervals each of length $2\beta^{n+1}$. Since $I_{n+1}^R = I_{n+1}^L + (1 - \beta)$, we have

$$I_{n+1}^L - I_{n+1}^R = (I_{n+1}^L - I_{n+1}^L) - (1 - \beta) = \beta(I_n - I_n) - (1 - \beta)$$

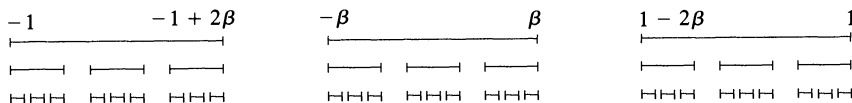
and similarly,

$$I_{n+1}^R - I_{n+1}^L = (I_{n+1}^R - I_{n+1}^R) + (1 - \beta) = \beta(I_n - I_n) + (1 - \beta)$$

so the induction hypothesis implies that both $I_{n+1}^L - I_{n+1}^R$ and $I_{n+1}^R - I_{n+1}^L$ are a disjoint union of 3^n closed intervals each of length $2\beta^{n+1}$. Now notice that

$$I_{n+1} - I_{n+1} = (I_{n+1}^L - I_{n+1}^R) \cup (I_{n+1}^R - I_{n+1}^L) \cup (I_{n+1}^L - I_{n+1}^L)$$

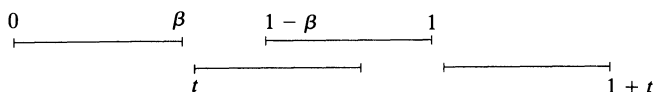
and, since $\beta < 1/3$, the unions are disjoint. So $I_{n+1} - I_{n+1}$ is a disjoint union of 3^{n+1} closed intervals each of length $2\beta^{n+1}$. The following picture illustrates $I_1 - I_1$, $I_2 - I_2$ and $I_3 - I_3$.



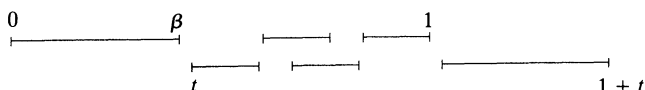
□

Exercise. As β increases from 0 to $1/3$, the sets $\Gamma_\alpha - \Gamma_\alpha$ grow in much the same way that the sets Γ_α grow as β increases from 0 to $1/2$. Show that if $\beta < 1/5$, then the difference set of the difference set $(\Gamma_\alpha - \Gamma_\alpha) - (\Gamma_\alpha - \Gamma_\alpha)$ is a Cantor set of measure zero and if $\beta \geq 1/5$, then $(\Gamma_\alpha - \Gamma_\alpha) - (\Gamma_\alpha - \Gamma_\alpha) = [-2, 2]$.

When $\beta \geq 1/3$ we know that $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is nonempty for all $t \in [-1, 1]$. Now we will show that for some values of β and for some $t \in (-1, 1)$, $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ will be as small as a nonempty set can be, i.e., a single point. Consider the following picture.



This picture represents I_1 and $I_1 + t$ for some, as yet, unspecified values of β and t . If we can choose β and t such that I_1^R and $I_1^L + t$ are positioned, relative to each other, the same way that I_0 and $I_0 + t$ are positioned, then, when we look at I_2^R and $I_2^L + t$, the above picture will reproduce itself, on a smaller scale, inside the intervals I_1^R and $I_1^L + t$. See the following picture (notice that the above picture, when it is reproduced below in the intervals I_1^R and $I_1^L + t$, has its orientation reversed).



So if β and t can be chosen so that the kind of self-similarity described above occurs, then for all n , I_n and $I_n + t$ will have only one pair of overlapping intervals. And then we will have only one point in $\Gamma_\alpha \cap (\Gamma_\alpha + t)$. The self-similarity we want can be described by looking at the proportion of $[0, 1]$ that lies to the left of $[t, 1 + t]$ and equating it to the proportion of $[1 - \beta, 1]$ that lies to the right of $[t, \beta + t]$. This can be expressed as

$$\frac{t}{1} = \frac{1 - (\beta + t)}{\beta}.$$

Solving for t we get

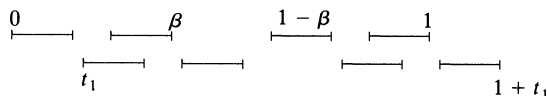
$$t = \frac{1 - \beta}{1 + \beta}. \quad (1)$$

But we also need to require that $\beta < t$. If we take the value of t given in equation (1) and put it in this last inequality, we get $\beta^2 + 2\beta - 1 < 0$ which is solved by $\beta < \sqrt{2} - 1$. So when $\beta < \sqrt{2} - 1$ and we give t the value $(1 - \beta)/(1 + \beta)$, then $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ will contain exactly one point. (Exercise: Find the value of this point.) When $\beta = \sqrt{2} - 1$, the intersection of Γ_α and $\Gamma_\alpha + t$ will be a countable number of endpoints along with the unique point that is in the interior of all the overlapping pairs of intervals. Notice that $\sqrt{2} - 1 > 1/3$.

For any $\beta \in (0, \sqrt{2} - 1)$ and any n , if we let

$$t_n = \beta^n \left(\frac{1 - \beta}{1 + \beta} \right),$$

then the intersection of Γ_α and $\Gamma_\alpha + t_n$ will contain exactly 2^n points. For example, consider the following picture, where $n = 1$.



This picture shows I_2 and $I_2 + t_1$, I_1^L and $I_1^L + t_1$ (and also I_1^R and $I_1^R + t_1$) are positioned, relative to each other, the same way that I_0 and $I_0 + t$ were in the last example. So the intersection of Γ_α and $\Gamma_\alpha + t_1$ will contain exactly two points.

These examples leave us with two questions. First, for $\beta \in [1/3, \sqrt{2} - 1]$ how often is the intersection of Γ_α and $\Gamma_\alpha + t$ a finite number of points? Second, for $\beta \geq \sqrt{2} - 1$, can the intersection of Γ_α and $\Gamma_\alpha + t$ be finite for any choice of $t \in (-1, 1)$? Before answering these questions, let's look at a simpler one; for $\beta \geq 1/3$, how often is the intersection of Γ_α and $\Gamma_\alpha + t$ a single point?

For any $\alpha \in (0, 1)$, let Λ_α denote the middle- α Cantor set defined in the interval $[-1, 1]$ (so $\Lambda_\alpha = 2\Gamma_\alpha - 1$). Λ_α can also be defined by

$$\Lambda_\alpha = \left\{ \sum_{i=0}^{\infty} u_i \beta^i (1 - \beta) \mid u_i \in \{-1, 1\} \text{ for all } i \right\}.$$

(What is the geometric significance for Λ_α of the *finite* sums $\sum_{i=0}^{n-1} u_i \beta^i (1 - \beta)$ where $u_i \in \{-1, 1\}$?) Notice that since Λ_α is a middle- α Cantor set, it has zero measure.

The next lemma shows that Λ_α contains all the points of $\Gamma_\alpha - \Gamma_\alpha$ for which $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is a single point. So we have an answer to our last question above. If $\beta \geq 1/3$, then for almost all $t \in [-1, 1]$, $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ contains more than one point.

Lemma 2. Choose $\beta \in (0, 1/2)$. If $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is a single point, then $t \in \Lambda_\alpha$.

Proof: Let x denote the single point in $\Gamma_\alpha \cap (\Gamma_\alpha + t)$. Let $\{s_0, s_1, s_2, \dots\}$ be the address in Γ_α of x and let $\{r_0, r_1, r_2, \dots\}$ be the address in Γ_α of $x - t$. Then

$$t = x - (x - t) = \sum_{i=0}^{\infty} (s_i - r_i) \beta^i (1 - \beta).$$

To show that $t \in \Lambda_\alpha$, we must show that $s_i - r_i \in \{-1, 1\}$ for all $i \geq 0$, i.e., we need to show that $s_i \neq r_i$ for all i . Suppose there is an i such that $s_i = r_i$. Let B and B' be the components of I_i that contain x and $x - t$ respectively. Assume that $s_i = r_i = 0$. Consider the following picture of $I_{i+1} \cap B$ and $(I_{i+1} \cap B') + t$.



(The two holes have length $\alpha\beta^i$ and the four closed intervals have length β^{i+1} .) Since $s_i = r_i = 0$, the point x is to the left of the open intervals (if $s_i = r_i = 1$, then the point x will be to the right of the open intervals). The two “halves” of the above picture are translations of each other, so in the right half of the picture the point $y = x + \beta^i(1 - \beta)$ will also be in $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ contradicting that x is the only point in this intersection. \square

Exercise. If $\Gamma_\alpha \cap (\Gamma_\alpha + t) = \{x\}$, then $t = 2x - 1$.

Let Λ_A denote the union of the images of Λ_α under all the affine maps of the form

$$T(x) = \beta^n x + \sum_{i=0}^{n-1} v_i \beta^i (1 - \beta)$$

where n is any integer greater than 0 and $v_i \in \{-1, 0, 1\}$ for $i = 0, \dots, n-1$. Then

$$\Lambda_A = \left\{ \sum_{i=0}^{\infty} s_i \beta^i (1 - \beta) \mid \text{for some } n > 0, s_i \in \{-1, 0, 1\} \text{ for } 0 \leq i \leq n-1 \right. \\ \left. \text{and } s_i \in \{-1, 1\} \text{ for } i \geq n \right\}.$$

Let's notice several things about Λ_A . First of all, $\Lambda_A \subset [-1, 1]$. In fact, $\Lambda_A \subset \Gamma_\alpha - \Gamma_\alpha$ (why?). Λ_A has measure zero since it is a countable union of sets of measure zero (the affine image of a set of measure zero has measure zero). The series representations of points in Λ_A are not unique. For example, if $\beta = \sqrt{2} - 1$ and $t = (1 - \beta)/(1 + \beta)$, then the sequences $\{0, 1, 1, 1, \dots\}$ and $\{1, -1, 1, -1, 1, -1, \dots\}$ can both be used to represent t .

The significance of Λ_A is that it contains all the points of $\Gamma_\alpha - \Gamma_\alpha$ for which $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is a finite set of points. This will be shown in the proof of the following theorem.

Theorem 3. If $\beta \geq 1/3$, then $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is a Cantor set for almost all $t \in [-1, 1]$.

Note. So $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ goes from being almost always empty when $\beta < 1/3$, to being almost always a Cantor set when $\beta \geq 1/3$. This shows that the change in the difference sets described by Theorem 1 is even more dramatic than was indicated by that theorem.

Proof: Let $\beta \in (0, 1/2)$. We will show that if $t \in (\Gamma_\alpha - \Gamma_\alpha) \setminus \Lambda_\alpha$, then $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is a Cantor set. The intersection of any two Cantor sets is a compact, totally disconnected set. To show that $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is a Cantor set, we only need to show that it is a perfect set.

Suppose that $t \in \Gamma_\alpha - \Gamma_\alpha$ and $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is not a perfect set, i.e., there is a point $x \in \Gamma_\alpha \cap (\Gamma_\alpha + t)$ and an $\varepsilon > 0$ such that $(x - \varepsilon, x + \varepsilon) \cap \Gamma_\alpha \cap (\Gamma_\alpha + t) = \{x\}$, or in other words, x is an isolated point in $\Gamma_\alpha \cap (\Gamma_\alpha + t)$. We'll show that $t \in \Lambda_A$. Choose n large enough so that $\beta^n < \varepsilon$. Let B denote the component of I_n that contains x and let B' be the component of I_n that contains $x - t$. Then $\Gamma_\alpha \cap B$ and $(\Gamma_\alpha \cap B') + t$ are two middle- α Cantor sets whose intersection contains only the point x . Let $\{s_0, \dots, s_{n-1}, 0, 0, 0, \dots\}$ be the address in Γ_α of the left hand endpoint of B and let $\{r_0, \dots, r_{n-1}, 0, 0, 0, \dots\}$ be the address in Γ_α of the left hand endpoint of B' . Let T denote the affine map

$$T(y) = \beta^{-n} \left(y - \sum_{i=0}^{n-1} s_i \beta^i (1 - \beta) \right).$$

The image of $\Gamma_\alpha \cap B$ under T is Γ_α . The image of $(\Gamma_\alpha \cap B') + t$ under T is

$\Gamma_\alpha + t'$ where

$$t' = T\left(t + \sum_{i=0}^{n-1} r_i \beta(1 - \beta)\right). \quad (2)$$

Because T is an affine map, we have $\Gamma_\alpha \cap (\Gamma_\alpha + t') = T((\Gamma_\alpha \cap B) \cap ((\Gamma_\alpha \cap B') + t))$, so $\Gamma_\alpha \cap (\Gamma_\alpha + t')$ contains only one point. Then the previous lemma implies that $t' \in \Lambda_\alpha$. By the definition of Λ_α , t' has a unique series representation of the form

$$t' = \sum_{i=0}^{\infty} u_i \beta^i (1 - \beta) \quad \text{where } u_i \in \{-1, 1\} \text{ for all } i. \quad (3)$$

Now solve for t in Equation (2) using the series (3) to represent t' . We get

$$\begin{aligned} t &= T^{-1}(t') - \sum_{i=0}^{n-1} r_i \beta^i (1 - \beta) \\ &= \beta^n t' + \sum_{i=0}^{n-1} (s_i - r_i) \beta^i (1 - \beta) \\ &= \beta^n \sum_{i=0}^{\infty} u_i \beta^i (1 - \beta) + \sum_{i=0}^{n-1} (s_i - r_i) \beta^i (1 - \beta) \\ &= \sum_{i=0}^{n-1} (s_i - r_i) \beta^i (1 - \beta) + \sum_{i=n}^{\infty} u_{i-n} \beta^i (1 - \beta) \end{aligned}$$

which implies that $t \in \Lambda_A$.

So far, we have only assumed that $\beta \in (0, 1/2)$ and we have shown that whenever $t \in (\Gamma_\alpha - \Gamma_\alpha) \setminus \Lambda_A$, then $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is a Cantor set. When $\beta \geq 1/3$, $\Gamma_\alpha - \Gamma_\alpha = [-1, 1]$. So when $\beta \geq 1/3$ and $t \in [-1, 1] \setminus \Lambda_A$, then $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is a Cantor set. The fact that Λ_A has measure zero completes the proof. \square

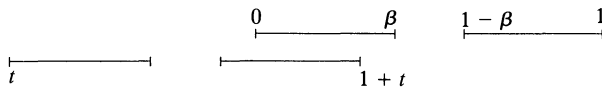
We have now answered one of our earlier questions. For $\beta \in [1/3, \sqrt{2} - 1]$, the intersection of Γ_α and $\Gamma_\alpha + t$ contains an infinite number of points for almost all t in $[-1, 1]$. The next theorem answers the other question.

Theorem 4. *If $\beta \in (\sqrt{2} - 1, 1/2)$, then $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ contains a Cantor set for all $t \in (-1, 1)$.*

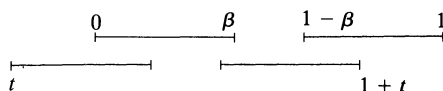
Proof: For $\beta \in (\sqrt{2} - 1, 1/2)$, define a function $f_\beta: [-1, 1] \rightarrow [-1, 1]$ by

$$f_\beta(t) = \begin{cases} \beta^{-1}t + (\beta^{-1} - 1) & \text{for } -1 \leq t \leq -\beta \\ \beta^{-1}t & \text{for } -\beta < t < \beta \\ \beta^{-1}t - (\beta^{-1} - 1) & \text{for } \beta \leq t \leq 1. \end{cases}$$

Let's see how f_β was derived. If $t \in [-1, -\beta]$, then I_1^L and $I_1^R + t$ overlap but $I_1^R + t$ and I_1^R do not overlap (except when $t = -\beta$ and they have a common endpoint).



If we concentrate only on the intervals I_1^L and $I_1^R + t$, then we can *renormalize* (or rescale) this pair of intervals by mapping I_1^L onto $[0, 1]$ and mapping $I_1^R + t$ onto an interval $[t', 1 + t']$ in such a way that the relation of $[0, 1]$ to $[t', 1 + t']$ is the same as the relation of I_1^L to $I_1^R + t$. The required value of t' is given by $f_\beta(t)$. If $t \in (-\beta, \beta)$, then I_1^L and $I_1^L + t$ overlap and of course I_1^R and $I_1^R + t$ also overlap (it may be that the pair I_1^L and $I_1^R + t$ or the pair I_1^R and $I_1^L + t$ overlap, but we will ignore these pairs in this case).



The value of $f_\beta(t)$ gives us, in this case, the renormalization of the pair of intervals I_1^L and $I_1^L + t$ (and simultaneously the renormalization of I_1^R and $I_1^R + t$). If $t \in [\beta, 1]$, then only I_1^R and $I_1^L + t$ overlap, and $f_\beta(t)$ renormalizes this pair of intervals.

The orbit $\{f_\beta^n(t)\}_{n=0}^\infty$ of the point t gives us information about $\Gamma_\alpha \cap (\Gamma_\alpha + t)$. For example, if the n th iterate of t is in the interval $(-\beta, \beta)$, then $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ contains at least two points (why?). If $f_\beta^n(t) \in (-\beta, \beta)$ for some finite set of integers $\{n_i\}_{i=1}^k$, then $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ contains at least 2^k points. What we will prove below is that if $f_\beta^{n_i}(t) \in (-\beta, \beta)$ for some infinite sequence of integers $\{n_i\}_{i=1}^\infty$, then $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ contains a Cantor set. But first let's show that for any $\beta \in (\sqrt{2} - 1, 1/2)$ and for any $t \in (-1, 1)$, there is a sequence $\{n_i\}_{i=1}^\infty$ of integers such that $f_\beta^{n_i}(t) \in (-\beta, \beta)$ for all $i \geq 1$.

To prove this fact, it suffices to show that if $t \in (-1, -\beta] \cup [\beta, 1)$, then for some n , $f_\beta^n(t) \in (-\beta, \beta)$ (so the orbit of t will always return to $(-\beta, \beta)$ after it leaves it). Notice that the points -1 and 1 are both fixed points for f_β . Suppose that $t \in [\beta, 1)$ (the proof for $t \in (-1, -\beta]$ is similar). The image under f_β of $[\beta, 1)$ is the interval $[2 - \beta^{-1}, 1)$ which, for $\beta > \sqrt{2} - 1$, is a subset of $(-\beta, 1)$. So the point $f_\beta(t)$ is contained in either $(-\beta, \beta)$ (in which case $n = 1$ and we are done) or it's still in $[\beta, 1)$. Notice that the distance of $f_\beta(t)$ from 1 , i.e., $1 - f_\beta(t) = 1 - (\beta^{-1}t - (\beta^{-1} - 1)) = \beta^{-1}(1 - t)$, is strictly greater than the distance that t was from 1 , i.e., $1 - t$ (since $\beta^{-1} > 1$). So as long as $f_\beta^n(t)$ stays in $[\beta, 1)$, its distance from 1 is $\beta^{-n}(1 - t)$. But eventually $\beta^{-n}(1 - t)$ will be strictly greater than $1 - \beta$ (the length of $[\beta, 1)$) and $f_\beta^n(t)$ will be in $(-\beta, \beta)$.

Now let's show that if $\beta \in (\sqrt{2} - 1, 1/2)$ and $t \in (-1, 1)$, then $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ contains a Cantor set. If B is a component of I_n , then B contains two components of I_{n+1} . Let $(B \cap I_{n+1})^L$ denote the left hand component of I_{n+1} contained in B and let $(B \cap I_{n+1})^R$ denote the right hand component. Let $B_0 = B'_0 = [0, 1]$. For $n \geq 0$, define inductively the following sequences of nested closed intervals

$$B_{n+1} = \begin{cases} (B_n \cap I_{n+1})^L & \text{if } f_\beta^n(t) \in [-1, \beta) \\ (B_n \cap I_{n+1})^R & \text{if } f_\beta^n(t) \in [\beta, 1] \end{cases}$$

and

$$B'_{n+1} = \begin{cases} (B'_n \cap I_{n+1})^R & \text{if } f_\beta^n(t) \in [-1, -\beta] \\ (B'_n \cap I_{n+1})^L & \text{if } f_\beta^n(t) \in (-\beta, 1]. \end{cases}$$

The lengths of the B_n go to zero, so $\bigcap_{n=0}^{\infty} B_n$ is a single point in Γ_α . Let x denote this point. Similarly, let $y = \bigcap_{n=0}^{\infty} B'_n$. B_n and B'_n have been defined so that B_n and $B'_n + t$ are the overlapping pair of components from I_n and $I_n + t$ that are renormalized by applying f_β to $f_\beta^{-1}(t)$. In the case where $f_\beta^{-1}(t) \in (-\beta, \beta)$, so f_β is simultaneously renormalizing two pairs of overlapping intervals, B_n and $B'_n + t$ are defined to be the left-hand pair. It follows that $\bigcap_{n=0}^{\infty} B_n = \bigcap_{n=0}^{\infty} (B'_n + t)$, i.e., that $x = y + t$. Let $\{s_0, s_1, s_2, \dots\}$ and $\{r_0, r_1, r_2, \dots\}$ denote the addresses in Γ_α of x and y respectively. So

$$s_n = \begin{cases} 0 & \text{if } f_\beta^n(t) \in [-1, \beta) \\ 1 & \text{if } f_\beta^n(t) \in [\beta, 1] \end{cases} \quad \text{and} \quad r_n = \begin{cases} 1 & \text{if } f_\beta^n(t) \in [-1, -\beta) \\ 0 & \text{if } f_\beta^n(t) \in (-\beta, 1] \end{cases}.$$

Notice that $s_n = r_n$ if and only if $f_\beta^n(t) \in (-\beta, \beta)$ and $s_n = r_n = 0$.

Let $\{n_i\}_{i=1}^{\infty}$ be the set of integers for which $f_\beta^{n_i}(t) \in (-\beta, \beta)$. Let \mathcal{A} denote the set of all sequences $\{a_0, a_1, a_2, \dots\} \in \{0, 1\}^{\mathbb{N}}$ such that $a_n = 0$ if $n \notin \{n_i\}_{i=1}^{\infty}$. If $a \in \mathcal{A}$, let

$$x_a = \sum_{n=0}^{\infty} (s_n + a_n) \beta^n (1 - \beta) \quad \text{and} \quad y_a = \sum_{n=0}^{\infty} (r_n + a_n) \beta^n (1 - \beta).$$

For any $a \in \mathcal{A}$, $x_a, y_a \in \Gamma_\alpha$ (recall that $s_n = r_n = 0$ if $n \in \{n_i\}_{i=1}^{\infty}$), $x_a - y_a = x - y = t$ and so $x_a \in \Gamma_\alpha \cap (\Gamma_\alpha + t)$. Let $\mathcal{X}_\mathcal{A} = \{x_a | a \in \mathcal{A}\}$. So $\mathcal{X}_\mathcal{A} \subset \Gamma_\alpha \cap (\Gamma_\alpha + t)$. Let's show that $\mathcal{X}_\mathcal{A}$ is a Cantor set. Since $\mathcal{X}_\mathcal{A} \subset \Gamma_\alpha$, it is totally disconnected. We need to show that it is closed and perfect. If $\{z_n\}$ is a sequence of points from $\mathcal{X}_\mathcal{A}$ that converge to z , then $z \in \Gamma_\alpha$ (since Γ_α is closed). It's not hard to show that for any M there is an N such that for $n > N$ the addresses of z_n and z will agree in the first M places. This implies that $z \in \mathcal{X}_\mathcal{A}$, so $\mathcal{X}_\mathcal{A}$ is closed. To show that $\mathcal{X}_\mathcal{A}$ is perfect, choose $a \in \mathcal{A}$ and define $a^k = \{a_0^k, a_1^k, a_2^k, \dots\} \in \mathcal{A}$ by $a_n^k = a_n$ if $n \leq k$ and $a_n^k = 0$ if $n > k$. Then for all k , $x_{a^k} \in \mathcal{X}_\mathcal{A}$ and the sequence $\{x_{a^k}\}_{k=0}^{\infty}$ converges to x_a . So $\mathcal{X}_\mathcal{A}$ is a perfect set. \square

Exercise. When $\beta < 1/3$, show that every $t \in \Gamma_\alpha - \Gamma_\alpha$ has a unique series representation

$$t = \sum_{i=0}^{\infty} s_i \beta^i (1 - \beta) \quad \text{where } s_i \in \{-1, 0, 1\}.$$

Call $\{s_0, s_1, s_2, \dots\}$ the address of t . Show that if $\{s_0, s_1, s_2, \dots\}$ is the address of a point t in $\Gamma_\alpha - \Gamma_\alpha$ and a is the cardinality of $\{n | s_n = 0\}$, then the cardinality of $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is equal to 2^a . (What about the case $\beta = 1/3$? See also [BM].)

Let's summarize what we have proven. If $\beta < 1/3$, then $\Gamma_\alpha - \Gamma_\alpha$ is a Cantor set with Lebesgue measure zero. If $\beta \in [1/3, 1/2)$, then $\Gamma_\alpha - \Gamma_\alpha = [-1, 1]$. In addition, if $\beta \in [1/3, \sqrt{2} - 1)$, then for almost all $t \in [-1, 1]$, $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is a Cantor set, but for some $t \in (-1, 1)$, $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is as small as one point. If $\beta = \sqrt{2} - 1$, then $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ is a Cantor set for all but a countably infinite subset of $[-1, 1]$, and for $t \in (-1, 1)$, the smallest cardinality that $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ can have is countably infinite. And if $\beta \in (\sqrt{2} - 1, 1/2)$, then $\Gamma_\alpha \cap (\Gamma_\alpha + t)$ contains a Cantor set for all $t \in (-1, 1)$. (The case where $\beta = \sqrt{2} - 1$ is actually an exercise for the reader.) These results can be generalized to arbitrary Cantor sets embedded in the real line if we replace the parameter β with something called the thickness of a Cantor set (see [K]).

In this article we have used the difference sets of middle- α Cantor sets as a way to make concrete the intuitive idea that a middle- α Cantor set with α close to zero

is larger than a middle- α Cantor set with α close to one. We did this using just cardinality and Lebesgue measure. Of course, what we lose by using such simple concepts of size is the ability to distinguish between, say, Γ_{α_1} and Γ_{α_2} when both α_1 and α_2 are in $(1/3, 1)$. For this, more sophisticated concepts such as Hausdorff or box-counting dimension are needed (see [E]).

REFERENCES

- [B] R. P. Boas, Jr., *A Primer of Real Functions*, Mathematical Association of America, Washington, D.C., 1966.
- [BM] N. C. Bose Majumder, *On the distance set of the Cantor middle third set, III*, this MONTHLY 72 (1965), 725–729.
- [E] G. A. Edgar, *Measure, Topology, and Fractal Geometry*, Springer-Verlag, New York, 1990.
- [HY] J. G. Hocking, G. S. Young, *Topology*, Addison-Wesley, Read, Mass., 1961, (reprinted by Dover Publications).
- [K] R. L. Kraft, *Intersections Of Thick Cantor Sets*, Mem. Amer. Math. Soc. **97** (1992), no. 468.

Department of Mathematics, Computer Science, and Statistics
Purdue University–Calumet
Hammond, IN 46323
roger@math.nwu.edu

Excerpt from “‘Billiards Is a Good Game’: Gamesmanship and America’s First Nobel Prize Scientist” by Norman Maclean, *The University of Chicago Magazine* 67 (Summer 1975, pp. 19–23).

For instance, Leonard Eugene Dickson, the outstanding mathematician, who at the time was writing his classic works on the theory of numbers, was sometimes a poor card player. Anton J. Carlson was also not a good bridge player, although he was nationally famous as an exponent of the scientific method of biological sciences.

...

Dickson, the master of numbers, was sometimes expectedly brilliant in a game where only 13×4 numbers were involved; his habitual troubles were at least partly environmental—he had come here by way of Texas. He almost consistently overbid and, when he lost three of four hands in a row, he would slam his cards down on the table and leave the room in a rage, always denouncing Carlson on the way out. No matter who had misplayed—Carlson, Michelson, or himself—he always denounced Carlson. While the cards were still shivering on the table he would shout, “Why the hell, Carlson, don’t you go back to your lab and feed your dogs? And don’t let Irene Castle catch you killing any of them.”

Overbidding three or four hands in a row and then blaming the great biologist seemed to put the great mathematician in the right state of mind to race back to his office and resume his classic studies on the theory of numbers.

Contributed by Jeffrey C. Lagarias
AT & T Bell Laboratories
600 Mountain Avenue, P.O. Box 636
Murray Hill, NJ 07974-0636

Morphisms, Squarefree Strings, and the Tower of Hanoi Puzzle

Jean-Paul Allouche, Dan Astoorian, Jim Randall,
and Jeffrey Shallit

1. INTRODUCTION. The *Tower of Hanoi* puzzle consists of three numbered pegs and N disks. Initially, the disks, which have radius $1, 2, \dots, N$, are all placed on peg 1 in increasing order of size, such that the smallest disk is on top, and the largest disk is at the bottom. See Figure 1.

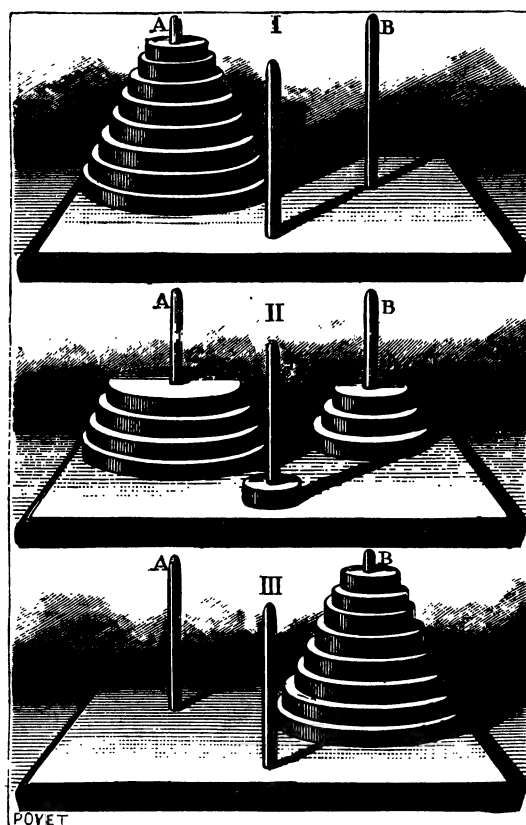


Figure 1. The Tower of Hanoi Puzzle

A *move* of the puzzle consists of taking the top disk off one peg and moving it to another. A move is called *legal* if it does not involve covering a smaller disk with a larger one. The object of the puzzle is to transfer all of the disks from peg 1 to

one of the other two pegs, using only legal moves. We are most interested in the *optimal* solutions to the puzzle; that is, solutions for N disks that use the *smallest* number of legal moves.

At each step there are six possible moves: moving a disk from peg 1 to peg 2; from peg 2 to peg 3; from peg 3 to peg 1; and their inverses. (Of course, not all of these may be legal at any given time.) We code these moves using the letters a , b , c and \bar{a} , \bar{b} , \bar{c} respectively.

Thus, for example, we can transfer 3 disks from peg 1 to peg 2 using the following sequence of moves:

$$a \bar{c} b a c \bar{b} a.$$

This sequence is optimal—no shorter sequence of legal moves will work. Actually, it is easy to construct an optimal solution for any number of disks; we discuss this further in the next section.

Fred Dixon, a student in an undergraduate course on formal languages taught by the fourth author, posed the following natural question:

When performing an optimal solution to the Tower of Hanoi with N disks, does there ever occur a sequence of moves that is immediately repeated?

In this article, we answer this question negatively. More precisely, we provide a self-contained proof, based on the theory of iterated morphisms, that optimal solutions to the puzzle are always *squarefree*. (More general results, including a discussion of the so-called “cyclic Tower of Hanoi”, will be given in [2].) We will use the Tower of Hanoi as our motivation to explore some concepts from formal language theory and combinatorics on words.

History. The Tower of Hanoi puzzle has been around for at least a hundred years, and is a favorite subject of recreational mathematicians. It was apparently invented by François-Édouard-Anatole Lucas (1842–1891), a famous French number theorist and recreational mathematician [14].

A nearly complete set of references to the puzzle may be obtained by consulting the articles of Allouche and Dress [4] and Hinz [12].

2. THE CLASSICAL OPTIMAL SOLUTION. We now describe the classical solution to the puzzle, which has the virtue of being optimal. To move N disks from peg 1 to peg 2, say, first move $N - 1$ disks (recursively) from peg 1 to peg 3, using peg 2 as intermediate storage. Once this is completed, move disk N from peg 1 to peg 2, and then finally move the $N - 1$ disks on peg 3 to peg 2 (recursively), using peg 1 as intermediate storage. Letting T_N be the number of moves in this solution, we see that $T_1 = 1$, and $T_N = 2T_{N-1} + 1$. By induction, one can easily prove that $T_N = 2^N - 1$.

That the classical solution is indeed optimal is not hard to see (although much fuss has been made on this point; see [10, 19]). To transfer N disks from one peg to another, we must at some point move disk N at least once. In order to move disk N , it must be alone on its peg, and some other peg must be empty; hence the remaining peg must contain the $N - 1$ smaller disks. Hence no algorithm can do better than to (i) move the first $N - 1$ disks to the same peg, using the optimal strategy; (ii) move disk N ; and (iii) move the $N - 1$ disks again, using the optimal strategy, and covering disk N . Hence, if $T'(N)$ denotes the total number of moves used by the optimal strategy to transfer N disks, we see that $T'(N) \geq 1 + 2T'(N - 1)$. Since $T'(1) = 1$, we see that $T'(N) \geq 2^N - 1$, as desired.

3. MORPHISMS. One of the most important concepts in formal language theory is that of the *homomorphism*, or just *morphism* for short. A morphism is a map h that assigns a string $h(t)$ to each symbol t of a finite alphabet Σ . This map is extended to Σ^* , the set of all finite strings of symbols chosen from Σ , using the rule $h(xy) = h(x)h(y)$.

For example, consider the map $\tau(0) = 01$ and $\tau(1) = 10$. Then $\tau(0110) = \tau(0)\tau(1)\tau(1)\tau(0) = 01101001$.

If the *length* $|h(\ell)|$ of the string $h(\ell)$ equals a constant k for all letters $\ell \in \Sigma$, then the morphism h is said to be *k-uniform*. A simple example of a 1-uniform morphism on the alphabet $\{a, b, c, \bar{a}, \bar{b}, \bar{c}\}$ is the morphism σ defined as follows:

$$\begin{aligned}\sigma(a) &= b & \sigma(\bar{a}) &= \bar{b} \\ \sigma(b) &= c & \sigma(\bar{b}) &= \bar{c} \\ \sigma(c) &= a & \sigma(\bar{c}) &= \bar{a}.\end{aligned}$$

This particular morphism has the following interpretation in terms of the Tower of Hanoi puzzle: if a string x has the effect of moving some disks from peg 1 to peg 2, using peg 3 as intermediate storage, then $\sigma(x)$ moves the same configuration of disks from peg 2 to peg 3, using peg 1 as intermediate storage. This morphism will play a major role in our results.

We define the morphism σ^{-1} in the obvious way, namely:

$$\begin{aligned}\sigma^{-1}(a) &= c & \sigma^{-1}(\bar{a}) &= \bar{c} \\ \sigma^{-1}(b) &= a & \sigma^{-1}(\bar{b}) &= \bar{a} \\ \sigma^{-1}(c) &= b & \sigma^{-1}(\bar{c}) &= \bar{b}.\end{aligned}$$

4. FIXED POINTS OF MORPHISMS. Consider the morphism $\tau(0) = 01$ and $\tau(1) = 10$ mentioned in Section 3. We can iterate such a map, using the definition $\tau^{n+1}(x) = \tau(\tau^n(x))$ for $n \geq 0$, and $\tau^0(x) = x$. In this case, it is easy to verify that the string $\tau^n(0)$ is a prefix of the string $\tau^{n+1}(0)$. Hence there is a *unique infinite sequence*

$$t = t_0 t_1 t_2 \cdots = 0110100110010110 \cdots$$

of which all of $\tau(0), \tau^2(0), \dots$ are prefixes. Since $\tau(t) = t$, we call t a *fixed point* of the map τ . (An annoying technicality is that we have only defined τ for finite strings, but the extension of τ to infinite sequences is easy to define.) There is one other fixed point \bar{t} of this map τ , obtained by replacing every 0 in t with a 1, and every 1 with a 0.

The infinite sequence t is the celebrated Prouhet–Thue–Morse sequence [16, 18, 15]. Our method for constructing t is a particular case of the following observation: if there is some symbol $a \in \Sigma$ such that $h(a) = ax$, where x is nonempty, then h has a unique fixed point p starting with a . If further h is *non-erasing* (i.e. $h(b)$ is not the empty string for any $b \in \Sigma$), then p is an infinite sequence. Note that the morphism τ that generates the Prouhet–Thue–Morse sequence is 2-uniform.

Cobham [7] was the first to study in detail the properties of uniform morphisms. See also [1, 9].

5. CODING THE SOLUTION TO THE TOWER OF HANOI. As we have seen in Section 2, for each integer $N \geq 0$, there exists a sequence of $2^N - 1$ moves that

constitutes an optimal solution to the Tower of Hanoi puzzle with N disks. Actually, there are *two* different solutions: one that results in the disks ending up on peg 2, and another that leaves all the disks on peg 3.

As above, we can code these solutions as strings of symbols, where each symbol represents a move. In what follows, we only consider the solution to the puzzle which takes the disks from peg 1 to peg 2 if N , the number of disks, is odd, and from peg 1 to peg 3 if N is even. This choice might at first seem unnatural, but its advantage is that the sequence of moves for $N + 1$ disks begins with the sequence of moves for N disks. Hence there is actually an *infinite* string of symbols

$$H = h_0 h_1 h_2 \cdots = a \bar{c} b a c \bar{b} a \bar{c} b \bar{a} c b a \bar{c} b \cdots$$

which codes the solution to the puzzle for $N = 1, 2, 3, \dots$ disks. Another interpretation is that H solves the puzzle for an *infinite* number of disks!

The infinite string H can be described as the limit of the sequence of strings $(H_i)_{i \geq 0}$, where each H_i is a string of length $2^i - 1$ that gives the solution to the puzzle for i disks. We can obtain a recursive formula for the H_i using the description of the optimal solution obtained previously:

Observation 1. We have $H_0 = \varepsilon$, the empty string, and

$$H_{2N+1} = H_{2N} a \sigma^{-1}(H_{2N}), \quad \text{for } N \geq 0; \quad (1)$$

$$H_{2N} = H_{2N-1} \bar{c} \sigma(H_{2N-1}), \quad \text{for } N \geq 1. \quad (2)$$

The recursion formulas may appear mysterious at first, but they are actually quite simple. For example, Eq. (2) says in order to solve the puzzle for $2N$ disks, first move $2N - 1$ disks from peg 1 to peg 2. Then, using the move \bar{c} , move the $2N$ th disk from peg 1 to peg 3. Finally, move $2N - 1$ disks from peg 2 to peg 3; note this is accomplished using the morphism σ that was introduced in Section 3.

Using these recursion formulas, we get, for example:

$$H_1 = a;$$

$$H_2 = a \bar{c} b;$$

$$H_3 = a \bar{c} b a c \bar{b} a;$$

The discerning reader will note that the sequence H does not appear to be periodic, but there is a kind of “pseudoperiodicity” of period 6. More precisely, we have the following lemma (where the notation x^∞ represents the infinite string $xxx \cdots$):

Lemma 2. We have

$$H = (a c b A c B)^\infty,$$

where A is a symbol denoting either a or \bar{a} , and similarly for the symbols B and c .

Proof: It follows easily by induction on N that

$$H_{2N+1} = (a c b A c B)^{(2^{2N+1}-2)/6} a, \text{ for } N \geq 0; \quad (3)$$

$$H_{2N} = (a c b A c B)^{(2^{2N}-4)/6} a c b, \text{ for } N \geq 1. \quad (4)$$

■

6. SQUAREFREE STRINGS. We say a finite string y is a *factor* of a (finite or infinite string) u if we can write $u = xyz$, where x is a finite string. We say a (finite or infinite) string u is *squarefree* (or *repeat-free*) if it has no factors of the form yy , where y is a finite, nonempty string. While it is easy to see that no string of length 4 or more over an alphabet of two letters can be squarefree, Thue produced an *infinite* squarefree string s over a three-letter alphabet.

Similarly, we say a (finite or infinite) string u is *cubefree* if it has no factors of the form yyy , where y is a finite, nonempty string. Thue showed that the infinite sequence $t = 01101001 \cdots$, discussed in Section 4, is cubefree. Actually, he showed even more: it is *overlap-free*. That is, it has no factors of the form $ayaya$, where a is a single letter.

History. Axel Thue initiated formal language theory in 1906 by studying squarefree and cubefree strings; see [18]. The book of Guy [11, Sect. E21] contains many references to papers on this topic. Also recommended is the paper of Bean, Ehrenfeucht, and McNulty [5] and the survey of Berstel [6].

7. DESCRIBING THE HANOI SEQUENCE H VIA MORPHISMS. In this section, we will prove the following surprising observation: we can describe the Hanoi sequence H as the fixed point of a certain 2-uniform morphism φ , defined below.

$$\begin{aligned}\varphi(a) &= a\bar{c} & \varphi(\bar{a}) &= ac \\ \varphi(b) &= c\bar{b} & \varphi(\bar{b}) &= cb \\ \varphi(c) &= b\bar{a} & \varphi(\bar{c}) &= ba.\end{aligned}$$

This fact was first noted in the paper of Allouche, Betrema, and Shallit [3]; also see Allouche and Dress [4].

Now it is readily observed that

$$\begin{aligned}\varphi(H_0) &= \varepsilon; \\ \varphi(H_1) &= a\bar{c}; \\ \varphi(H_2) &= a\bar{c}bac\bar{b}; \\ \varphi(H_3) &= a\bar{c}bac\bar{b}a\bar{c}b\bar{a}cb\bar{a}\bar{c};\end{aligned}$$

so we see that $\varphi(H_0)a = H_1$, $\varphi(H_1)b = H_2$, $\varphi(H_2)a = H_3$, and $\varphi(H_3)b = H_4$. We might reasonably guess that this pattern repeats, and that $\varphi(H_n)a = H_{n+1}$ for n even, and $\varphi(H_n)b = H_{n+1}$ for n odd.

To prove this guess, we first prove a lemma:

Lemma 3. *Let $\Sigma = \{a, b, c, \bar{a}, \bar{b}, \bar{c}\}$. Then for $w \in \Sigma^*$, we have*

$$\varphi(\sigma(w)) = \sigma^{-1}(\varphi(w)); \quad (5)$$

$$\varphi(\sigma^{-1}(w)) = \sigma(\varphi(w)). \quad (6)$$

Proof: It suffices to verify the two equations for each letter in Σ . ■

We can now prove our guess about the form of $\varphi(H_i)$:

Lemma 4. *For all $i \geq 0$, we have $\varphi(H_{2i})a = H_{2i+1}$ and $\varphi(H_{2i+1})b = H_{2i+2}$.*

Proof: By induction on i . The assertions are easily verified for $i = 0$. Then for all $i > 0$,

$$\begin{aligned}
 \varphi(H_{2i})a &= \varphi(H_{2i-1} \bar{c} \sigma(H_{2i-1})) a \quad (\text{by (2)}) \\
 &= \varphi(H_{2i-1}) ba \varphi(\sigma(H_{2i-1})) a \\
 &= \varphi(H_{2i-1}) ba \sigma^{-1}(\varphi(H_{2i-1})b) \quad (\text{by (5)}) \\
 &= H_{2i} a \sigma^{-1}(H_{2i}) \quad (\text{by induction}) \\
 &= H_{2i+1}. \quad (\text{by (1)})
 \end{aligned}$$

The proof that $\varphi(H_{2i+1})b = \varphi(H_{2i+2})$ is similar and is left to the reader. This completes the proof of Lemma 4. ■

As a corollary, we have

Corollary 5. *The Tower of Hanoi sequence H is a fixed point of the map φ ; i.e. $H = \varphi(H)$.*

Proof: It suffices to give an infinite sequence of prefixes of H that are mapped by φ into longer prefixes of H . But this follows from Lemma 4, for we see that $\varphi(H_{2i})$ is a prefix of length $2^{2i+1} - 2$ of H_{2i+1} ; hence a prefix of H . A similar statement holds for $\varphi(H_{2i+1})$. ■

8. THE HANOI SEQUENCE H IS SQUAREFREE. We are now ready to tackle the problem of showing that the Tower of Hanoi sequence H is squarefree. The proof we will present is due to the second and third authors, who were undergraduates at the University of Waterloo.

We will need three lemmas. The first two lemmas, given in [4], follow easily from Lemma 2:

Lemma 6. *Ignoring the bars, the symbols in H are periodic of period 3.*

This lemma can also be found in [17].

Lemma 7

$$h_i = \begin{cases} a, & \text{if } i \equiv 0 \pmod{6}; \\ b, & \text{if } i \equiv 2 \pmod{6}; \\ c, & \text{if } i \equiv 4 \pmod{6}. \end{cases}$$

Hence the symbols appearing in even-numbered positions in H are unbarred.

This lemma (essentially) can be found in [8].

Lemma 8. *H does not contain four consecutive unbarred symbols.*

Proof: For if it did have four consecutive unbarred symbols, say $h_i h_{i+1} h_{i+2} h_{i+3}$, then i is either even or odd. If i is even, say $i = 2k$, then since H is a fixed point of φ , we can write $\varphi(h_k h_{k+1}) = h_i h_{i+1} h_{i+2} h_{i+3}$. But since only a barred symbol has an image under φ that consists of unbarred symbols, h_k and h_{k+1} must both be barred, contradicting Lemma 7.

If i is odd, then by Lemma 7, h_{i-1} must be unbarred, and we can repeat the same argument on $h_{i-1}h_ih_{i+1}h_{i+2}$. ■

Here is the main theorem of our paper:

Theorem 9. *The Tower of Hanoi sequence H contains no factor of the form xx , for x a nonempty string.*

Proof: Suppose H did contain a “square” factor

$$xx = x_1x_2 \cdots x_nx_1x_2 \cdots x_n.$$

Then without loss of generality, we may assume that xx is (i) as short as the shortest square occurring in H and (ii) among all squares with this length, xx occurs “earliest” in H .

We now divide the proof into three cases: (A) $|x|$ is odd; (B) $|x|$ is even, and begins at an even-numbered position in H , and (C) $|x|$ is even, and begins at an odd-numbered position in H .

Case (A). If $n = |x|$ is odd, then it is easy to see from Lemma 7 that all the symbols of x must be unbarred. (For example, if $xx = h_jh_{j+1} \cdots h_{j+n-1}h_{j+n}h_{j+n+1} \cdots h_{j+2n-1}$ for j even, then $h_j = x_1$, $h_{j+2} = x_3, \dots, h_{j+n-1} = x_n$ are all unbarred, and $h_{j+n+1} = x_2$, $h_{j+n+3} = x_4, \dots, h_{j+2n-2} = x_{n-1}$ are also unbarred. A similar argument applies if j is odd.) This contradicts Lemma 8 if $n \geq 3$, and if $n = 1$, it contradicts Lemma 6.

Case (B). Suppose $n = |x|$ is even, say $n = 2m$, and that our supposed square xx begins at an even-numbered position in H . More precisely, assume that

$$xx = h_{2j}h_{2j+1} \cdots h_{2j+2m-1}h_{2j+2m}h_{2j+2m+1} \cdots h_{2j+4m-1}.$$

Then since H is a fixed point of φ , we have

$$\varphi(h_jh_{j+1} \cdots h_{j+m-1}h_{j+m}h_{j+m+1} \cdots h_{j+2m-1}) = xx.$$

Since each letter has a distinct image under φ , we must have $h_{j+m} = h_j$, $h_{j+m+1} = h_{j+1}$, \dots , $h_{j+2m-1} = h_{j+m-1}$. Thus $h_jh_{j+1} \cdots h_{j+2m-1} = yy$, where $y = h_jh_{j+1} \cdots h_{j+m-1}$. But then yy is a nonempty square that is shorter than xx , a contradiction.

Case (C). Suppose $n = |x|$ is even, and that xx begins at an odd-numbered position in H . Thus we can write

$$qxx = qh_{2j+1} \cdots h_{2j+n}h_{2j+n+1} \cdots h_{2j+2n}$$

for some symbol $q = h_{2j}$. Now by Lemma 6, n must be divisible by 3; hence $n \equiv 0 \pmod{6}$. Then by Lemma 7, $q = h_{2j+n}$. Hence $q = x_n$. Thus

$$x_nx_1x_2 \cdots x_{n-1}x_nx_1x_2 \cdots x_{n-1} = h_{2j} \cdots h_{2j+2n-1}$$

is a square of the same length as xx , but occurring earlier in H , a contradiction.

This completes the proof of our theorem. ■

9. CONCLUSIONS. We have given a brief tour of the theory of iterated morphisms and squarefree strings, and used this theory to prove that no sequence of moves is ever immediately repeated when performing the optimal solution to the Tower of Hanoi puzzle.

There are many interesting topics we could not explore further in this short article. For example, the fixed points of k -uniform homomorphisms are strongly related to the so-called “automatic sequences,” which have many interesting number-theoretic properties. The interested reader is referred to the survey of Allouche [1].

The study of squarefree and cubefree sequences properly belongs to the field of “combinatorics on words”; for further information, the reader may want to study the book of Lothaire [13].

ACKNOWLEDGMENTS. Part of this work was done while the first author was visiting the University of Waterloo.

We would like to thank the referee, Anna Lubiw, and Bill Smyth for their suggestions.

REFERENCES

1. J.-P. Allouche. Automates finis en théorie des nombres. *Expo. Math.* **5** (1987), 239–266.
2. J.-P. Allouche, D. Astoorian, and J. Shallit. The Towers of Hanoi and k th-power-free words. In preparation.
3. J.-P. Allouche, J. Betrema, and J. Shallit. Sur des points fixes de morphismes du monoïde libre. *RAIRO Informatique Théorique* **23** (1989), 235–249.
4. J.-P. Allouche and F. Dress. Tours de Hanoi et automates. *RAIRO Informatique Théorique et Applications* **24** (1990), 1–15.
5. D. A. Bean, G. Ehrenfeucht, and G. McNulty. Avoidable patterns in strings of symbols. *Pacific J. Math.* **85** (1979), 261–294.
6. J. Berstel. Some recent results on squarefree words. In M. Fontet and K. Mehlhorn, editors, *STACS '84*, Vol. 166 of *Lecture Notes in Computer Science*, pages 14–25. Springer-Verlag, 1984.
7. A. Cobham. Uniform tag sequences. *Math. Systems Theory* **6** (1972), 164–192.
8. P. Cull and C. Gerety. Is Towers of Hanoi really hard? *Congr. Numer.* **47** (1985), 237–242.
9. F. M. Dekking, M. Mendès France, and A. J. van der Poorten. Folds! *Math. Intelligencer* **4** (1982), 130–138, 173–181, 190–195.
10. J. S. Frame and B. M. Stewart. Solution of problems 3918. *Amer. Math. Monthly* **48** (1941), 216–219.
11. R. K. Guy. *Unsolved Problems in Number Theory*, Vol. I of *Unsolved Problems in Intuitive Mathematics*. Springer-Verlag, New York, 1981.
12. A. M. Hinz. The tower of Hanoi. *Enseign. Math.* **35** (1989), 289–321.
13. M. Lothaire. *Combinatorics on Words*, Vol. 17 of *Encyclopedia of Mathematics and Its Applications*. Addison-Wesley, 1983.
14. E. Lucas. Le calcul et les machines à calculer. *Assoc. Française pour l'Avancement des Sciences; Comptes Rendus* **13** (1884), 111–141.
15. M. Morse. Recurrent geodesics on a surface of negative curvature. *Trans. Amer. Math. Soc.* **22** (1921), 84–100.
16. E. Prouhet. Mémoire sur quelques relations entre les puissances des nombres. *C. R. Acad. Sci. Paris* **33** (1851), 225.
17. R. S. Scorer, P. M. Grundy, and C. A. B. Smith. Some binary games. *Math. Gazette* **28** (1944), 96–103.
18. A. Thue. Über unendliche Zeichenreihen. *Norske vid. Selsk. Skr. I. Mat. Nat. Kl. Christiana* **7** (1906), 1–22. Reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, 1977.
19. D. Wood. The towers of Brahma and Hanoi revisited. *J. Recreational Math.* **14** (1981), 17–24.

Allouche:

C.N.R.S., L.M.D.

Luminy Case 930

F-13288, Marseille,

Cedex 9, FRANCE

allouche@lmd.univ-mrs.fr

Randall:

Waterloo, Ontario, CANADA

jrandall@cs.utoronto.ca

Astoorian:

Mississauga, Ontario, CANADA

djast@utopia.druid.com

Shallit:

Department of Computer Science

University of Waterloo

Waterloo, Ontario, N2L 3G1, CANADA

shallit@graceland.uwaterloo.ca

NOTES

Edited by: John Duncan

Sierpinski's Theorem is Deducible from Euler and Dirichlet

A. A. Ageev

A famous theorem of Dirichlet says that the sequence

$$an + b, \quad n = 1, 2, \dots$$

contains infinitely many primes if a and b are relatively prime integers. It is a similar long-standing conjecture that the quadratic sequence

$$n^2 + t, \quad n = 1, 2, \dots$$

contains an infinite number of primes for any positive integer t . The first result in this direction is due to Sierpinski [6] who showed that for any M there exists t' such that the sequence

$$n^2 + t', \quad n = 1, 2, \dots$$

contains at least M primes. Recently several new proofs and extensions of Sierpinski's theorem have appeared [4][3][1].

In this note we show that a slightly stronger result can be easily derived from the following well-known number theory facts:

Theorem 1 (Euler) ([5, Theorem 251]) *Every prime of the form $4n + 1$ is representable as a sum of two squares. This representation is unique up to the order of the summands.*

Theorem 2 (Dirichlet) ([2, p. 34]) *The series*

$$\sum_{\text{prime } p \equiv b \pmod{a}} 1/p$$

where a, b are relatively prime integers, diverges.

Denote by P the set of primes of the form $4n + 1$. Note that by Theorem 1 the set of primes in $\{(2k + 1)^2 + 4l^2 | k, l \in \mathbb{Z}_+\}$ coincides with P . For any $k \in \mathbb{Z}_+$, denote by P_k the set of primes in $\{4l^2 + 4k(k + 1) + 1 | l \in \mathbb{Z}_+\}$. Now Theorem 1 can be reformulated in the following way:

$$\bigcup_{k \in \mathbb{Z}_+} P_k = P, \tag{1}$$

$$P_{k'} \cap P_{k''} = \emptyset \quad \text{for all } k', k'' \in \mathbb{Z}_+ \text{ such that } k' \neq k''. \tag{2}$$

Theorem 3. For any $k_0, M \geq 0$ and any $\delta \in [0, 1[$ there exists a positive integer $k^* \geq k_0$ such that

$$|P_{k^*}| > M(2k^* + 1)^\delta.$$

(To obtain Sierpinski’s theorem, take $k_0 = \delta = 0$.)

Proof: By (1) and (2) we may formally write

$$\sum_{p \in P} 1/p = \sum_{k \in \mathbb{Z}_+} \sum_{p \in P_k} 1/p. \tag{3}$$

Note first that

$$\sum_{p \in P_k} 1/p < \sum_{l \in \mathbb{Z}_+} \frac{1}{4l^2 + 1} < +\infty \quad \forall k \in \mathbb{Z}_+. \tag{4}$$

Now assume to the contrary that, for some $M, k_0 \geq 0$ and some $\delta \in [0, 1[$,

$$|P_k| \leq M(2k + 1)^\delta \quad \forall k \in \mathbb{Z}_+ \cap [k_0, +\infty[.$$

This implies that, for any $k \in \mathbb{Z}_+ \cap [k_0, +\infty[$,

$$\sum_{p \in P_k} 1/p \leq \frac{|P_k|}{4k(k + 1) + 1} \leq \frac{M}{(2k + 1)^{2-\delta}} \leq \frac{M}{(k + 1)^{2-\delta}}. \tag{5}$$

It follows from (4) and (5) that the series in the right hand side of (3) converges and thereby so does the series in the left hand side. But this contradicts the statement of Theorem 2. ■

The following generalization is in fact an easy consequence of Theorem 3.

Corollary 1. For any $k_0, M, r, s \geq 0$ and any $\delta \in [0, 1[$ there exists a positive integer $k^* \geq k_0$ such that

$$|P_{k^*}| > M(rk^* + s)^\delta. \tag{6} \quad \blacksquare$$

REFERENCES

1. U. Abel and H. Siebert, *Sequences with large numbers of prime values*, Amer. Math. Monthly 100 (1993), 167–169.
2. H. Davenport, *Multiplicative number theory*, 2nd ed., Springer-Verlag, 1980.
3. R. Forman, *Sequences with many primes*, Amer. Math. Monthly 99 (1992), 548–557.
4. B. Garrison, *Polynomials with large numbers of prime values*, Amer. Math. Monthly 97 (1990), 316–317.
5. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, 1938.
6. W. Sierpinski, *Les binômes $x^2 + n$ et les nombres premiers*, Bull. Soc. Royale Sciences Liege 33 (1964), 259–260.

*Institute of Mathematics
Universitetskii pr. 4
Novosibirsk 630090
Russia
ageev@math.nsk.su*

On Nonnegativity of Symmetric Polynomials

F. Matúš

The nonnegativity of elementary symmetric polynomials (cf. [5])

$$x_1 + x_2 + \cdots + x_n, x_1x_2 + x_1x_3 + \cdots + x_{n-1}x_n, \dots, x_1x_2 \cdots x_n$$

implies the nonnegativity of all x_1, x_2, \dots, x_n in the domain of real numbers, see [3], Problem 10C. For complex numbers this ceases to be valid and thus to get the same conclusion we must assume more. The nonnegativity of the polynomials below in all powers $k \geq 1$ of the arguments

$$x_1^k + x_2^k + \cdots + x_n^k, x_1^kx_2^k + x_1^kx_3^k + \cdots + x_{n-1}^kx_n^k, \dots, x_1^kx_2^k \cdots x_n^k$$

is also not yet enough. Namely, the choice of all n -th roots of unity, $n \geq 3$ prime, renders $x_1^k, x_2^k, \dots, x_n^k$ to be the all roots or ones and having in mind the Viète formulae for roots of unity we see immediately the claimed nonnegativity.

Where $\alpha = (\alpha_1, \dots, \alpha_n)$ is an n -tuple of nonnegative integers and $x = (x_1, \dots, x_n)$ is an n -tuple of complex numbers, we shall write $x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}$. The symmetric polynomials (called in [6] “Potenzproduktsummen”)

$$p_{[\alpha]}(x_1, \dots, x_n) = p_{[\alpha]}(x) = \sum_{\beta \in [\alpha]} x^\beta, \quad x \in \mathbb{C}^n,$$

where β ranges over the set $[\alpha]$ of all permutations of α , include all the polynomials displayed above. For them α is to be taken as $k\iota_M$ where M is a nonempty subset of $N = \{1, 2, \dots, n\}$ and ι_M is the indicator function of M . A natural question asks which of these polynomials must be supposed nonnegative in order to conclude that all the arguments are nonnegative. More formally, denoting by \mathcal{A} the family of all our n -tuples α , find all $\mathcal{B} \subset \mathcal{A}$ such that $x \in \mathbb{C}^n$ and $p_{[\alpha]}(x) \geq 0$, $\alpha \in \mathcal{B}$, imply $x_i \geq 0$, $i \in N$.

The aim of this short note is first to show that any \mathcal{B} with finite complement in \mathcal{A} will suffice. Then we explore the same question replacing the complex numbers by square complex matrices with either the usual matrix product under commutativity assumption or the coordinatewise Hadamard product. At the same time the nonnegativity is considered either in all coordinates at once or as positive semidefiniteness.

Theorem. *Let $x \in \mathbb{C}^N$ and let the values $p_{[\alpha]}(x)$ be nonnegative for all $\alpha \in \mathcal{A}$. Then necessarily all coordinates x_i , $i \in N$, are nonnegative.*

Proof: We proceed by induction on the cardinality n of the set N . The case $n = 1$ is trivial. Let $n \geq 2$ and $x \in \mathbb{C}^N$, having at least one nonzero coordinate, fulfill the assumptions of the theorem. If the support $M = \{i \in N; x_i \neq 0\} \neq \emptyset$ of x is a proper subset of N then the restriction $y = x|_M \in \mathbb{C}^M$ satisfies $p_{[\alpha|_M]}(y) = p_{[\alpha]}(x)$ for all $\alpha \in \mathcal{A}$ such that $\alpha|_{N-M} = 0$, and then the induction argument yields $x_i = y_i > 0$, $i \in M$. So we now suppose that $\prod_{i \in N} x_i \neq 0$.

Let us suppose first that we can partition the set N into two nonempty subsets I and J such that $|x_i| > |x_j|$ for $i \in I$ and $j \in J$. Where b is an upper bound of the set $\{|x_i|; i \in N\}$ and an n -tuple α is zero on J , we can estimate the difference

$$\begin{aligned} & \left| p_{[\alpha+k\iota_I]}(x) \prod_{i \in I} x_i^{-k} - p_{[\alpha|_I]}(x|_I) \right| \\ & \leq \sum \left\{ |x^\beta| \prod_{i \in I} |x_i|^{-k}; \beta \in [\alpha + k\iota_I], \exists j \in J: \beta_j > 0 \right\} \\ & \leq n! b^{\sum_{i \in N} \alpha_i} \sum \left\{ \prod_{i' \in I'} |x_{i'}|^k \prod_{i \in I} |x_i|^{-k}; I' \subset N, I' \neq I, |I'| = |I| \right\}, \quad k \geq 1. \end{aligned}$$

The point is, due to the choice of the partition, that

$$p_{[\alpha|_I]}(x|_I) = \lim_{k \rightarrow \infty} \frac{p_{[\alpha+k\iota_I]}(x)}{x^{k\iota_I}}.$$

However, x^{ι_I} is a positive number since the numbers $x_i, i \in N$, being roots of the polynomial equation $\sum_{l=0}^n (-1)^l c_l x^{n-l} = 0$ with the nonnegative coefficients $c_l = p_{[\alpha]}(x)$ (α 's are indicators), can be either nonnegative or match into pairs of mutually adjoint complex numbers. This enables effective use of the induction argument to establish $x_i > 0, i \in I$. A little reflexion about the vector $y = (x_i^{-1})_{i \in N}$ and the polynomial values

$$p_{[\alpha]}(y) = \sum_{\beta \in [\alpha]} \frac{1}{x^\beta} = \sum_{\beta \in [k\iota_N - \alpha]} \frac{x^\beta}{x^{k\iota_N}} = \frac{p_{[k\iota_N - \alpha]}(x)}{x^{k\iota_N}} \geq 0$$

(k is taken sufficiently high) persuades us that the above reasoning, applied to y , gives $x_j > 0, j \in J$.

It remains to analyse the case $|x_i| = a, i \in N$. Due to homogeneity of the polynomials there is no loss of generality in assuming $a = 1$. Denoting by \mathcal{A}_m the family $\{\alpha \in \mathcal{A}; \alpha_i \leq m, i \in N\}$ we can write

$$q_m(x) = \sum_{[\alpha] \subset \mathcal{A}_m} p_{[\alpha]}(x) = \sum_{[\alpha] \subset \mathcal{A}_m} \sum_{\beta \in [\alpha]} x^\beta = \sum_{\alpha \in \mathcal{A}_m} \prod_{i \in N} x_i^{\alpha_i} = \prod_{i \in N} \sum_{l=0}^m x_i^l, \quad m \geq 0.$$

Now, no $x_i, i \in N$, can be a root of unity different from 1 for $x_i^m = 1$ would imply $q_{km-1}(x) = 0, k \geq 1$, and this yields $p_{[\alpha]}(x) = 0$ for every $\alpha \in \mathcal{A}$, contradicting $\prod_{i \in N} x_i = 1$. If one of the coordinates x_i is, however, not a root of unity then an old number-theoretic fact says the sequence $(x_i^k, k \geq 1)$ is dense in the unit circle and, by compactness, for some numbers $k_1 < k_2 < \dots$ every sequence $(x_j^{k_l}, l \geq 1)$ converges to some $y_j, j \in N$. The limit point y_i can be adjusted to -1 . The continuity of the polynomials implies that $y = (y_j)_{j \in N}$ satisfies the nonnegativity assumptions, which contradicts the presence of -1 among its coordinates. It follows that $x_i = 1, i \in N$. ■

Remarks. 1. When the word “nonnegative” is replaced in the formulation of the theorem by “positive” the new claim becomes, obviously, valid. Let us mention that we encountered the need for these assertions during an analysis of special symmetric random arrays.

2. The assumptions of the theorem can be further weakened. For example, the nonnegativity of the polynomials can be required only for $\alpha \in \mathcal{B} = \mathcal{A} - \mathcal{A}_m$ (apply the theorem to $(x_i^{m+1})_{i \in N}$ and $(x_i^{m+2})_{i \in N}$, consecutively).

Corollary. Let X_1, X_2, \dots, X_n be normal and pairwise commuting square complex matrices. The polynomials $p_{[\alpha]}(X_1, \dots, X_n)$, $\alpha \in \mathcal{A}$, are positively semidefinite if and only if all the matrices X_1, \dots, X_n possess this property.

Proof: By [2] (Ch. IX, Par. 15) there exists a unitary matrix U (UU^* is the identity matrix) such that all matrices UX_iU^* , $i \in N$, are diagonal. Then $p_{[\alpha]}(X_1, \dots, X_n) \geq 0$ (a self-explaining abbreviation for positive semidefiniteness), $\alpha \in \mathcal{A}$, if and only if

$$Up_{[\alpha]}(X_1, \dots, X_n)U^* = p_{[\alpha]}(UX_1U^*, \dots, UX_nU^*) \geq 0, \quad \alpha \in \mathcal{A},$$

and by the theorem this is equivalent to the nonnegativity of all elements of the diagonal matrices UX_iU^* , $i \in N$, i.e., to $UX_iU^* \geq 0$ and finally to $X_i \geq 0$, $i \in N$. ■

Remarks. 1. The assumption of normality cannot be omitted as is easily seen from the example of the following two commuting matrices

$$X_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad X_2 = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}.$$

Note that any symmetric polynomial of X_1 and X_2 is a multiple of the identity matrix.

2. If we want to drop the assumption of commutativity we must redefine the symmetric polynomials $p_{[\alpha]}$. No reasonable generalization of the corollary can be expected as there exist two matrices

$$X_1 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \geq 0 \quad X_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \geq 0.$$

such that the Hermitian matrix $X_1X_2 + X_2X_1$ is not positively semidefinite (see [1]).

3. If the matrices of $X = (X_i)_{i \in N}$ are nonnegative (coordinatewisely) then trivially also all polynomials $p_{[\alpha]}(X)$ are nonnegative but the contrary is not valid as the following matrices

$$X_1 = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad X_2 = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

testify (X_1 and X_2 are idempotent and they commute with $X_1X_2 = 0$).

4. Let us consider the symmetric polynomials $p_{[\alpha]}(X_1, \dots, X_n)$ built as above but with Hadamard product on the place of the usual one (see [4]). Since this product preserves positive semidefiniteness once matrices $X = (X_i)_{i \in N}$ have this property all their polynomials $p_{[\alpha]}(X)$ share it, too. The converse is, however, not the case: the polynomials $p_{[\alpha]}$ in the two matrices

$$X_1 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad X_2 = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

are positively semidefinite.

REFERENCES

1. N. N. Chan and M. K. Kwong, Hermitian matrix inequalities and a conjecture, *Amer. Math. Monthly* 92 (1985) 533–541.
2. F. R. Gantmacher, *The Theory of Matrices*. Chelsea Publishing Company, New York 1959.
3. P. R. Halmos, *Problems for Mathematicians Young and Old*. Dolciani Mathematical Expositions 12, The Mathematical Association of America 1991.
4. R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, Cambridge 1986.

5. A. Kurosh, *Higher Algebra*. Mir Publishers, Moscow 1975.
6. O. Perron, *Algebra, Vol. 1: Die Grundlagen*. Walter de Gruyter & Co., Berlin W 10 and Leipzig, 1932.

*Institute of Information Theory and Automation
Academy of the Sciences of Czech Republic
Pod Vodárenskou věží 4, 182 08 Prague
Czech Republic
matus@utia.cas.cz*

New Tricks for Old Trees: Maps and the Pigeonhole Principle

N. Graham, R. C. Entringer, and L. A. Székely

A TRICKY PROOF. The second author conjectured and the first author proved the following theorem on spanning trees of the hypercube Q_n :

Theorem 1. *For every spanning tree T of the hypercube Q_n , there exists an edge e in $E(Q_n) - E(T)$ whose addition to T forms a cycle of length at least $2n$.*

Proof: Let T be any spanning tree in Q_n . For any vertex v of Q_n there is an antipodal vertex \bar{v} in Q_n . There is a unique path in T from v to \bar{v} ; orient its first edge towards \bar{v} and repeat this process for every vertex v . Since the tree T has fewer edges than vertices, the pigeonhole principle implies there is an edge (u, v) which has received two orientations. The distance between u and \bar{u} (and between v and \bar{v}) is n in Q_n and hence is at least n in T . Finally, if (u, v) is an edge in Q_n then (\bar{u}, \bar{v}) also is an edge in Q_n . Hence (u, v) and (\bar{u}, \bar{v}) and the edges of the $u - \bar{u}$ and $v - \bar{v}$ paths yield the cycle sought. ■

Theorem 1 is clearly best possible since adding any further edge to any breadth-first search tree T yields a cycle of length $2n$ or less. Theorem 1 also implies the result that any spanning tree T of Q_n has a diameter which is at least $2n - 1$ [6].

After introducing some definitions we show that the proof technique for Theorem 1 applies to several tree properties. In a connected graph G we denote by $d(x, y)$ the distance between vertices x and y .

Definition. The *eccentricity* of $v \in V(G)$ is given by $\text{ecc}(v) = \max_{u \in V(G)} d(u, v)$.

Definition. The graph G has an *antipodal isomorphism* if

- (i) $\text{ecc}(v)$ is the diameter of G for every $v \in V(G)$,
- (ii) for every $v \in V(G)$ there exists a unique $\bar{v} \in V(G)$ such that $d(v, \bar{v})$ is the diameter of G , and
- (iii) the map $\phi: V(G) \rightarrow V(G)$ defined by $\phi(v) = \bar{v}$ is an isomorphism of G .

Theorem 1'. Assume the graph G has an antipodal isomorphism. Then, for every spanning tree T of G , there exists an edge e in $E(G) - E(T)$ whose addition to T forms a cycle of length at least two times the diameter of G . ■

Some other examples of graphs admitting an antipodal automorphism are the one-dimensional skeletons of regular and semi-regular polytopes which admit a central symmetry [4] and toroidal grids (of which the hypercube is a special case).

MORE THAN A TRICK. We soon realized that the method of the proof of Theorem 1 applies to a number of “little theorems” or exercises, yielding shorter proofs than the usual ones. The standard proofs for most of these exercises can be found in Lovász [7, Chapter 6, pp. 41–42].

Definition. The *center* of a tree T is the set of vertices ν in $V(T)$ with minimum eccentricity $\text{ecc}(\nu)$.

Theorem 2 (Jordan [8]). *The center of a tree T contains either one vertex or two adjacent vertices.*

Proof: For every vertex ν find a vertex ν' such that $d(\nu, \nu') = \text{ecc}(\nu)$. Orient the first edge of the unique $\nu - \nu'$ path towards ν' . By the pigeonhole principle, some edge (p, q) receives two orientations. Observe that any vertex $\nu \neq p$ in the component with p in $T - (p, q)$ has $\text{ecc}(\nu) > \text{ecc}(p)$ and any vertex $\nu \neq q$ in the component with q in $T - (p, q)$ has $\text{ecc}(\nu) > \text{ecc}(q)$. Assume, without loss of generality, that $\text{ecc}(p) \geq \text{ecc}(q)$. If $\text{ecc}(p) > \text{ecc}(q)$, then q alone realizes the minimum eccentricity. Otherwise $\text{ecc}(p) = \text{ecc}(q)$ and the second possibility in the theorem holds. ■

Definition. The maximal subtrees containing a vertex ν of a tree T as an endvertex are called the *branches* of T at ν . The *weight* of a branch B , denoted by $\text{bw}(B)$, is the number of edges in it. The *branch weight* of a vertex ν , $\text{bw}(\nu)$, is the maximum weight of a branch at ν . The *centroid* of a tree T is the set of vertices ν of T with minimum branch weight $\text{bw}(\nu)$.

Theorem 3 (Jordan [8]). *The centroid of a tree T contains either one vertex or two adjacent vertices.*

Proof: For every vertex ν find a branch B_ν of maximum weight at the vertex ν . Orient the only edge of B_ν incident with ν towards B_ν . The rest of the proof is identical with that of Theorem 2, changing ecc to bw . ■

Definition. An *endomorphism* of a tree T is a map $\phi: V(T) \rightarrow V(T)$ with the property that for any $(p, q) \in E(T)$ either $\phi(p) = \phi(q)$ or $(\phi(p), \phi(q)) \in E(T)$.

Theorem 4 (Lovász [7]). *Any endomorphism ϕ of a tree T has either a fixed vertex or a fixed edge.*

Proof: Assume ϕ has no fixed vertex. Then, for every vertex ν , there is a unique non-trivial path in T from ν to $\phi(\nu)$. Orient the first edge of this path towards $\phi(\nu)$. Repeating this for every vertex ν implies, by the pigeonhole principle, the existence of an edge (p, q) with two orientations. Hence the edge (p, q) separates

p from $\phi(p)$ and q from $\phi(q)$. Since ϕ is an endomorphism, the only possibility is $\phi(p) = q$ and $\phi(q) = p$, i.e., (p, q) is a fixed edge. ■

Theorem 5 (Clarke [3, Theorem 3], Caro and Schönheim [5, Theorem 2]). *A tree T has a perfect matching if and only if deleting any vertex v yields exactly one odd component.*

Proof: The condition for the perfect matching is necessary since, deleting v , the component containing the mate of v in the matching is odd and no other component is odd. The condition is sufficient for the following reason. For any vertex v , there is a unique edge e joining v and the odd component of $T - v$; orient this edge e toward the odd component. Denote the other end vertex of e by v' . Observe that in $T - v'$ the component containing v is odd (since T has an even number of vertices). Therefore, repeating the same procedure with v' instead of v , we obtain e with the opposite orientation. Repeating the procedure for all vertices, we obtain a set of edges with two orientations which cover every vertex exactly once, i.e., a perfect matching.

We note that this proof also shows that if a tree has a perfect matching then the matching is unique. This fact is implicit in results of Beineke and Plummer [2; Theorem 2].

Theorem 6. For a simple graph G on $n \geq 2$ vertices, the following propositions are equivalent:

- (i) G is connected and has $n - 1$ edges
- (ii) G is cycle-free and has $n - 1$ edges
- (iii) G is cycle-free and connected
- (iv) G is connected but deleting any edge disconnects G
- (v) any two vertices of G are connected by a unique path
- (vi) G is cycle-free and adding any new edge creates a cycle.

It is not our intention here to give a complete proof. The cyclic proof (in this cyclic order) may use standard proofs except for the following two.

(i) \rightarrow (ii) Assume that G has a cycle C . Put a cyclic orientation on the edges of C . For any vertex not in C , there exists a shortest path connecting v to C since G is connected. For each such vertex v choose a shortest path and orient toward C the (unique) edge incident with v on this path. Observe that from each vertex not on C an oriented edge goes out and no edge is oriented in both directions. Hence the number of edges is at least n , a contradiction.

(vi) \rightarrow (i) G is connected, otherwise, adding an edge between two vertices in different components would not create a cycle. Fix any vertex z of G and for each $x \in V(G)$, $x \neq z$, orient toward z the edge incident with x in the unique xz path in G . Since G is acyclic, every edge is oriented exactly once; thus G has $n - 1$ edges. (Any edge e of G is the last edge of a path starting at z , since G is connected. This is the unique path connecting its end vertices, since G is cycle-free. Therefore e gets an orientation. No edge gets two orientations since this would yield a cycle.) ■

There is an interesting alternative proof that (vi) \rightarrow (i). After noting, as before, that G is connected, we assume G has m vertices and prove $m = n - 1$ with the following argument. Define the map $\tau: E(\overline{G}) \rightarrow [E(G)]^2$, where \overline{G} is the complement of G , as follows: adding any edge $e \in E(\overline{G})$ creates a unique cycle, let $\tau(e)$

be the set of two edges adjacent to e in this cycle. Now the reader can show that τ is both an injection and a surjection and, so, a bijection. Thus $\binom{n}{2} - m = \binom{m}{2}$ and, hence, $m = n - 1$.

ACKNOWLEDGMENT. We thank a referee for providing important additional references.

REFERENCES

1. C. Berge, "Graphs," 2nd ed., North-Holland, Amsterdam, 1985.
2. L. W. Beineke and M. D. Plummer, On the 1-factors of a non-separable graph, *J. Combin. Theory* 2 (1967) 285–289.
3. F. H. Clarke, A graph polynomial and its applications, *Discrete Math.* 3 (1972) 305–313.
4. H. S. M. Coxeter, "Regular Polytopes," 2nd ed., MacMillan, N.Y., 1963.
5. Y. Caro and J. Schönheim, Generalized 1-factorization of trees, *Discrete Math.* 33 (1981) 319–321.
6. N. Graham and F. Harary, Changing and unchanging the diameter of a hypercube, *Discrete Appl. Math.* 37–38 (1992) 265–274.
7. L. Lovász, "Combinatorial Problems and Exercises," Akadémiai Kiadó, Budapest, and North-Holland, Amsterdam, 1979.
8. C. Jordan, Sur les assemblages de lignes, *J. Reine Angew. Math.* 70 (1869) 185–190.

Graham:
Computing Research Laboratory
New Mexico State University
Las Cruces, NM 88003

Entringer:
Department of Mathematics
University of New Mexico
Albuquerque, NM 87131

Székely:
Department of Computer Science
Eötvös Loránd University
Budapest, HUNGARY

The constructs of the mathematical mind are at the same time free and necessary. The individual mathematician feels free to define his notions and set up his axioms as he pleases. But the question is will he get his fellow mathematician interested in the constructs of his imagination. We cannot help the feeling that certain mathematical structures which have evolved through the combined efforts of the mathematical community bear the stamp of a necessity not affected by the accidents of their historical birth. Everybody who looks at the spectacle of modern algebra will be struck by this complementarity of freedom and necessity.

—H. Weyl, 1951

THE COMPUTER SCIENCE SAMPLER

Edited by: Catherine C. McGeoch

Do You Know the Way to Vertex A ?

Jeffrey Ondich

When I want to communicate with a member of my wife's family, it is often most efficient to send my message through my mother-in-law, Elinor. At any given moment, Elinor knows where everyone is, what they are doing, and the easiest way to reach them. Especially if I want to broadcast a message to the whole family, Elinor provides me with a fast, reliable communication mechanism.

When I send electronic mail to a friend on the Internet, things are a bit different. My message gets divided into one or more *packets* (depending on how long-winded I am that day), and each packet is handed from machine to machine until it reaches my friend's computer. There is no central, omniscient authority like Elinor on the Internet, so each machine along each packet's path needs to make its own *routing decisions*. That is, each machine must decide which of its neighbor machines should receive my packet to ensure the most efficient delivery of my message.

Elinor provides my family's communication system with centralized control. Of course, if Elinor gets sick or heads out in the camper with no forwarding address, our system will be in trouble. To avoid similar problems, the Internet relies on distributed control of its communications. In this column, I will describe one class of distributed algorithms used by portions of the Internet to make routing decisions.

THE SET-UP. Most large computer networks are divided into smaller sub-networks, or *autonomous systems*. All communications between distinct autonomous systems pass through machines called *gateways*. A gateway is a computer (often but not always specially designed for communications duty) that is part of at least two autonomous systems. One gateway can pass a packet directly to another only if both gateways are a part of the same autonomous system.

A graph is a natural way of representing a computer network. For our purposes, the vertices will be gateways, and two vertices will be connected by an edge if the two corresponding gateways are part of the same autonomous system. Each edge will have a positive length, representing the cost of passing a packet from one gateway to the next across the autonomous system they share. The path a packet follows through our network will correspond to a walk on this graph.

And now, a bit of notation.

- d_{vw} = the length of the edge connecting vertices v and w ($d_{vv} = 0$).
- D_{vw} = the length of the shortest path between vertices v and w .
- H_{vw}^k = the length of the shortest path from v to w involving no more than k edges or *hops*.
- $N(v)$ = the set of v 's neighbors (that is, v and the vertices adjacent to v).
- $D(v, w)(t)$ = an estimate of D_{vw} stored at v at time t . (When the time is not important, we will use $D(v, w)$.)
- $F(v, w)(t)$ = the first vertex on what v believes to be the shortest route from v to w at time t .

THE BELLMAN-FORD ALGORITHM. Suppose every gateway in our network contains a copy of the network's gateway graph. When a packet addressed to gateway A arrives at gateway B , B must decide which of its neighbor machines should receive the packet next. If B has been using its time wisely, it has already computed a shortest path to every other vertex on the graph, including A , and will know the proper neighbor to hand the packet to.

One algorithm for computing the lengths of the shortest paths between vertices is called the *Bellman-Ford algorithm* [1]. It is an iterative algorithm that computes every H_{vw}^k during its k th iteration, like so:

$$\begin{aligned} H_{vv}^k &= 0 \\ H_{vw}^0 &= +\infty \\ H_{vw}^k &= \min_{u \in N(v)} \{d_{vu} + H_{uw}^{k-1}\}, \end{aligned} \quad (1)$$

where v and w are distinct vertices, and k is any positive integer. The idea behind (1) is this: vertex v can send a packet to w in at most k hops by handing the packet to vertex u and instructing u to get the packet to w in at most $k - 1$ hops. The length of the resulting trip will be $d_{vu} + H_{uw}^{k-1}$. By taking the minimum of these distances over all of v 's neighbors, we find the shortest k -hop distance from v to w .

The Bellman-Ford algorithm ends when $H_{vw}^k = H_{vw}^{k-1}$ for all v and w . At that point, each H_{vw}^k will be equal to the correct distance D_{vw} . These distances satisfy *Bellman's Equation*:

$$\begin{aligned} D_{vv} &= 0 \\ D_{vw} &= \min_{u \in N(v)} \{d_{vu} + D_{uw}\} \end{aligned} \quad (2)$$

A proof that $H_{vw}^k = D_{vw}$ for sufficiently large k depends upon the fact that (2) has a unique solution, consistent of the shortest path distances D_{vw} . We will use this uniqueness later.

THE DISTRIBUTED BELLMAN-FORD ALGORITHM. To apply the Bellman-Ford algorithm to a graph, you need to know the whole graph. But to assume that every gateway in a network has a complete and accurate picture of the whole network is to assume too much. Computers crash and come back on line, and edge lengths between adjacent gateways change (if, for example, lengths are dependent on the amount of packet traffic). It takes time for word of these changes to make

its way through the network. So gateways need to make their decisions based on their most current information, which may or may not be correct.

Routing decisions are really buck-passing decisions. The decider does not need to know the whole path to the destination—just the first step of that path. A distributed variant of the Bellman-Ford algorithm provides a mechanism for maintaining this first-step information at each vertex.

Each gateway on a network using the distributed Bellman-Ford algorithm stores a *distance vector* for the network, a table whose entries consist of the identity of a destination, an estimate of the distance to that destination, and the first gateway along the shortest route to that destination. For example, the distance vector stored at *B* in Figure 1 would be

Destination <i>w</i>	Distance <i>D(B, w)</i>	First Step <i>F(B, w)</i>
A	2	E
B	0	B
C	2	D
D	1	D
E	1	E

When a packet arrives at a gateway, the gateway consults its distance vector entry for the packet’s destination, hands the packet to the appropriate neighbor, and goes about its business. (If the gateway sees no valid route to the destination, it might return the packet to its sender, store the packet in a dead or delayed letter file, or just discard the packet.) If every gateway’s distance vector contains the correct distances (that is, $D(v, w) = D_{vw}$ for all v and w) and best first-steps, then this method delivers packets along the shortest paths.

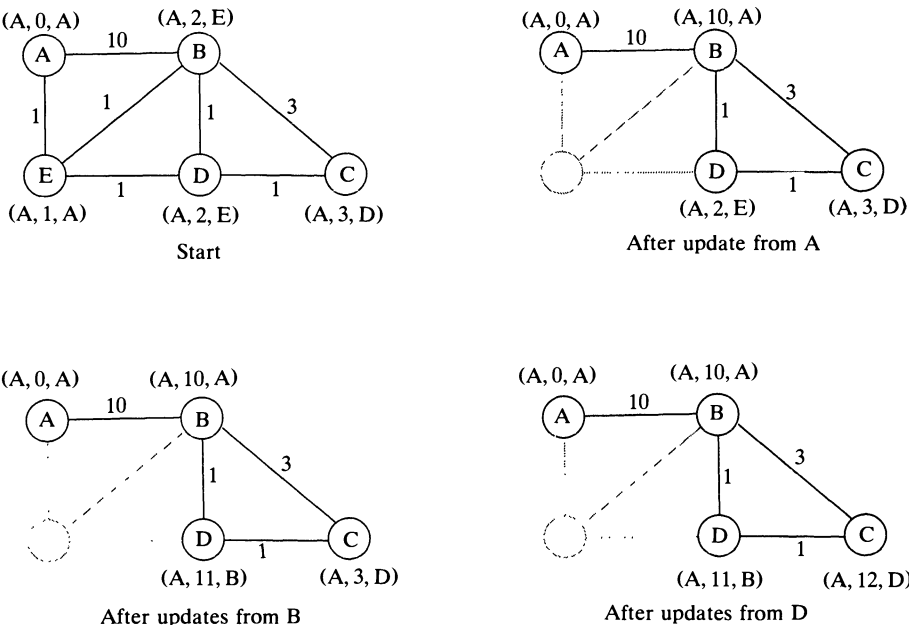


Figure 1

Since networks change over time, we need a way of changing distance vectors. The way this is normally done is to insist that each gateway periodically send an *update*—that is, a copy of its distance vector—to each of its neighbors.

Suppose A receives an update from B . Then for each destination C , A *accepts* (instead of ignoring) the update by setting

$$\begin{aligned} D(A, C) &= D(B, C) + d_{AB} \\ F(A, C) &= B \end{aligned}$$

if any of the following is true:

1. $D(B, C) + d_{AB} < D(A, C)$
2. $F(A, C) = B$
3. $F(A, C)$ has not sent an update to A for a long time, and is assumed to be unreachable. (In [2], a “long time” is defined as 180 seconds.)

Let’s see how this works in an example. Figure 1 shows distance vector entries for destination A . These entries were correct before gateway E crashed. After E has been out of commission for a while, B and D mark their routes to A invalid, since these routes first pass through E . C ’s route to A is also invalid, but C doesn’t know it yet.

Suppose now that (1) A sends an update to B , then (2) B sends updates to C and D , and finally, (3) D sends updates to B and C . Figure 1 shows the modified distance vector entries after each update.

The distance vector entries for destination A are all correct after the three updates have been sent. Of course, the order in which the updates are processed is important. Had D sent its update to C before receiving the update from B , then C ’s distance vector would not be correct yet. But as long as every gateway keeps sending updates occasionally, the distance vectors will eventually converge to correct values. The following proposition formalizes this claim.

You may find it odd that the hypotheses of this proposition require the network to remain static. The point of the updating scheme is, after all, to allow distance vectors to adapt to changes in the network. You should imagine a sequence of network changes separated by periods of no change. After a change, the distance estimates can have essentially arbitrary values. Hypothesis (a) allows arbitrary initial values, and so the proposition adequately models a changing network.

Proposition. *Let N be a finite, connected network that does not change after time t_0 (no gateways or lines crash or come up, and no edge lengths change). Suppose that for any pair of distinct vertices v and w , and any time $t \geq t_0$: (a) $D(v, w)(t_0)$ is an arbitrary element of $[0, +\infty]$; (b) either $F(v, w)(t_0)$ is a vertex in N , or the route from v to w is marked as invalid; (c) $D(v, v)(t) = 0$; and (d) the sequence of times at which v receives updates from each neighbor w is unbounded. Then there exists a finite time after which $D(v, w)(t) = D_{vw}$ for all v, w , and t .*

Proof: Start by defining H_{vw}^k as above, and L_{vw}^k by

$$\begin{aligned} L_{vv}^k &= 0 \\ L_{vw}^0 &= 0 \\ L_{vw}^k &= \min_{u \in N(v)} \{d_{vu} + L_{uw}^{k-1}\}, \end{aligned} \tag{3}$$

where v and w are distinct vertices, and k is any positive integer.

L_{vw}^k and H_{vw}^k are estimates of D_{vw} computed using the Bellman-Ford algorithm with the lowest and highest possible initial conditions. By using (1), (2), and (3) to make four simple and nearly identical induction arguments on k , we can show

$$L_{vw}^k \leq L_{vw}^{k+1} \leq D_{vw} \leq H_{vw}^{k+1} \leq H_{vw}^k$$

for every v, w , and k . Furthermore, both H_{vw}^k and L_{vw}^k converge to D_{vw} in a finite number of steps. To see this for L_{vw}^k , note that the number of possible values for L_{vw}^k is finite. Since L_{vw}^k is non-decreasing and bounded above, $L_{vw}^k = L_{vw}^{k-1}$ for sufficiently large k . Thus, for large enough k , $L_{vv}^k = 0$, and

$$L_{vw}^k = L_{vw}^{k+1} = \min_{u \in N(v)} \{d_{vu} + L_{uw}^k\}$$

for each v and w . That is, the numbers L_{vw}^k (for this sufficiently large fixed k) satisfy (2). By the uniqueness of the solution of (2), we have $L_{vw}^k = D_{vw}$, as desired. A similar argument shows that H_{vw}^k converges to D_{vw} (see [1]).

Now that we have two sequences squeezing in upon D_{vw} , we will fit our distance vector estimates $D(v, w)(t)$ between L_{vw}^k and H_{vw}^k so the estimates will also converge in finite time to D_{vw} .

More specifically, for every integer $k \geq 0$ there exists a $t_k > 0$ such that whenever $t \geq t_k$,

$$L_{vw}^k \leq D(v, w)(t) \leq H_{vw}^k \quad (4)$$

for every pair of vertices v and w .

Our induction on k to prove the left inequality of (4) starts easily. Since $L_{vw}^0 = 0$, $D(v, w)(t_0) \geq 0$, and $d_{vw} > 0$ for every v and w , the left half of (4) holds for $k = 0$.

Now suppose (4) holds for a fixed k . Hypothesis (d) guarantees that there is a time $t_{k+1} > t_k$ such that each vertex receives at least one update from each of its neighbors between times t_k and t_{k+1} . During this time interval, hypothesis (b) ensures that every vertex v must accept at least one update for each destination w . Since

$$L_{vw}^{k+1} = \min_{u \in N(v)} \{d_{vu} + L_{uw}^k\}$$

and $D(u, w)(t) \geq L_{uw}^k$ for every $t \geq t_k$, we have

$$D(v, w)(t) \geq L_{vw}^{k+1}$$

for any $t \geq t_{k+1}$. By induction then, the left half of (4) holds for all k . The proof of the right half of (4) is a bit trickier than this, and can, again, be found in [1].

Assuming the claims for H_{vw}^k are true, we now know that for sufficiently large k and $t \geq t_k$, $L_{vw}^k = D_{vw} = H_{vw}^k$, and $L_{vw}^k \leq D(v, w)(t) \leq H_{vw}^k$. That is, there is a finite time after which $D(v, w)(t) = D_{vw}$ for every v, w , and t . \square

THE REAL WORLD. The preceding proposition guarantees that the distance vectors will adapt to network changes within a finite time, which should satisfy most mathematicians. In practice, of course, some finite times—the short ones—are better than others.

Consider Figure 1 again, but suppose that just the AE edge goes down, while E and its other connections continue to function. Vertices B, C, D and E will send out about 10 updates each before their A -distances inch up above 10 (try it, supposing updates go out in alphabetical order). If updates are sent from each gateway every 30 seconds (the maximum allowed in [2]), service to A will be disrupted for at least 5 minutes. In the meantime, packets bound for A will circle

around the network until the distance vectors are correct, further clogging the system.

If A had crashed instead of just the AE connection, then the $D(v, A)$'s would climb as before, but this time without bound. Real systems use this *counting to infinity* behavior to identify unreachable gateways.

FURTHER READING. You can find a detailed and more general discussion of the Bellman-Ford algorithm in [1], which contains a wealth of ideas about the analysis of computer networks. For a pragmatic approach, [2] discusses implementation issues in distance vector algorithms, and defines the Routing Information Protocol (RIP), a formal specification of a distance vector routing scheme.

For more on routing algorithms, see [3]. For the entertaining story of a major network crisis caused by some subtle routing algorithm flaws, see [5] and [4].

References [2] and [5] are *Requests for Comments*. You can get *RFC's* by anonymous ftp at nic.merit.edu, in the documents/rfc directory. If you have a stout heart and a real interest in the workings of the Internet, *RFC's* are indispensable reading.

REFERENCES

1. D. Bertsekas and R. Gallager, *Data Networks*, Prentice Hall, 1992.
2. C. Hedrick, Routing Information Protocol, RFC 1058, June 1988.
3. R. Perlman, Routing Protocols, *Internet System Handbook*, D. Lynch and M. Rose, eds., Addison Wesley, 1993.
4. R. Perlman, Fault-Tolerant Broadcast of Routing Information, *Computer Networks* (7), 1983, pp. 395–405.
5. E. Rosen, Vulnerabilities of Network Control Protocols: An Example, RFC 789, 1982.

Department of Mathematics and Computer Science
Carleton College
One North College Street
Northfield MN 55057-4025
jondich@carleton.edu

Standard mathematics has recently been rendered obsolete by the discovery that for years we have been writing the numeral five backward. This has led to reevaluation of counting as a method of getting from one to ten. Students are taught advanced concepts of Boolean algebra, and formerly unsolvable equations are dealt with by threats of reprisals.

—Woody Allen

Return to Mathematical Circles by Howard W. Eves,
Boston: Prindle, Weber and Schmidt, 1988.

THE EVOLUTION OF . . .

Edited by Abe Shenitzer

Mathematics, York University, North York, Ontario M3J 1P3, Canada

On the Calculus of Variations and Its Major Influences on the Mathematics of the First Half of Our Century.* Part I.

Erwin Kreyszig

1. JOHANN BERNOULLI'S BRACHYSTOCHRONE. CARATHÉODORY'S METHOD. The calculus of variations evolved from the differential and integral calculus, "the calculus," for short. An initial motivation of the latter was the determination of extrema of *functions*, as shown by the title of the earliest relevant published paper *A new method for the determination of maxima and minima . . .* (Leibniz, 1684). However, the calculus had to be extended to the calculus of variations in order to take care of more general problems involving the determination of stationary values of *functionals*,[†] given in the simplest case by a definite integral involving an unknown function and boundary conditions. The earliest (not very simple) *solved* problem of this kind was Newton's determination of the shape of a gun shell of least air resistance (letter to Gregory of July 14, 1694).

The earliest problem that received general publicity [due to a rather bombastic advertisement in *Acta Eruditorum* by Johann Bernoulli (1667–1748)] in 1696 was the problem of determining the *brachystochrone*, the curve along which a particle will fall from one given point to another in the shortest time. This problem was solved by Newton, Leibniz, and Johann Bernoulli as well as by his brother Jacob (1654–1705), the solution being a *cycloid*. Thus 1696 can be called the birthyear of the calculus of variations. Johann Bernoulli not only posed that problem but also gave a solution capable of extensive generalization worked out in 1908 by Carathéodory. The resulting general method was later named after Carathéodory.

2. SIMPLEST GENERAL PROBLEMS. Although they arose from different geometric and physical applications, many of the early problems led to functionals that depended on real functions defined on an interval and satisfying boundary condi-

*Abbreviated version of a paper with the same title.

[†]A functional is a function defined on a set of functions. A stationary value of a functional is its value at a "point" (= function) that satisfies the necessary conditions for an extremum (see Section 3 below).

tions, and all functionals were of the *same form* (as needed to create a general theory).

$$J[y] = \int_{x_0}^{x_1} L(x, y, y') dx, \quad y(x_0) = y_0, \quad y(x_1) = y_1, \quad x_0 < x_1. \quad (2.1)$$

The task was to determine a function $y(x)$ that satisfied the boundary conditions in (2.1) and rendered $J[y]$ stationary, possibly yielding a minimum or a maximum of $J[y]$. At this early stage, the existence and uniqueness of solutions was “obvious” for physical reasons because a solution could be verified experimentally if desired. Also, with the concept of function not yet sharply defined, nobody made an attempt to characterize the set of functions in which such a $y(x)$ was to be found. This was accomplished well over one hundred years later in the works of Jacobi (1804–51) around 1835 and, especially, of Weierstrass (1815–97) around 1880.

3. EULER AND LAGRANGE. As the birthyear of the *theory* of the calculus of variations one usually considers 1744, the year in which Euler published his famous book *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes, sive solutio problematis isoperimetrici latissimo sensu accepti* (A method for discovering curved lines that enjoy a maximum or minimum property, or the solution of the isoperimetric problem taken in its widest sense). Thus Euler replaced “art of invention” (*ars inveniendi*), a very popular term in the works of Tschirnhaus and in other works of Leibniz’s time, by “method of invention,” a remarkable turn toward systematization. This book, a landmark in the development of the subject, contained the *Euler equation*

$$\frac{\partial L}{\partial y} - \frac{d}{dx} \left(\frac{\partial L}{\partial y'} \right) = 0, \quad (3.1)$$

(first published by Euler in 1736) as a necessary condition for $y(x)$ satisfying (2.1) to yield a minimum of $J[y]$. In more explicit form it is the equation

$$L_{y'y'} y'' + L_{y'y} y' + L_{y'x} - L_y = 0. \quad (3.2)$$

This equation suggests calling (2.1) a *regular problem* when $L_{y'y'}$ is never zero, and then assuming that $L_{y'y'} > 0$.

Euler’s book also contains a fascinating collection of 66 problems. Carathéodory, the editor of the book as a volume of Euler’s *Works*, said that it

“is one of the most beautiful mathematical works ever written. We cannot emphasize enough the extent to which that *Lehrbuch* over and over again served later generations as a prototype in the endeavor of presenting special mathematical material in its [logical, intrinsic] connection.”

Euler’s inspiration came from geometry and even more from the *principle of least action*, according to which nature realizes all motions in the most economical manner; more precisely, among all possible ways of reaching a given goal, nature chooses the one which minimizes the *action integral* $\int m v ds$ over the path (m = mass, v = speed, s = arc length). The beginning of the principle is often dated back to Leibniz because of a (lost) letter he is supposed to have written on the principle in 1707, but the question is still an open one. The principle is usually named after de Maupertuis (1698–1759), president of the Berlin Academy under Frederick the Great. Actually, Euler most likely discovered it earlier, formulated it

mathematically more rigorously, and applied it to a nontrivial problem (involving central forces). In contrast to this, Maupertuis published the principle (in 1744 and 1746) in a vague and almost theological form. He defended vigorously his (questionable) priority, but failed to realize that a rigorization of the principle would call for specification of conditions to be satisfied by the motions with which the actual motion was to be compared. Accordingly, his main merit seems to be that he was *searching for* a minimum principle.

The great significance of the calculus of variations in mathematical physics is due to the transparent and coordinate-free form that the laws of nature take in this calculus. That fact became apparent in the work of Euler, and even more impressively in that of Lagrange. In his path-breaking memoir *Essai d'une nouvelle méthode pour déterminer les maxima et les minima des formules intégrales indéfinies* (1760–61) Lagrange substantially overtook Euler (as Euler was well aware). His work was a milestone in the development of the field and in its application to geometry and analytical mechanics. In it he invented the “method of variations” together with the symbol δ . His new idea was to use “*comparison functions*,”

$$\bar{y} = y + \varepsilon \eta, \quad \eta \in C^2([x_0, x_1]), \quad \eta(x_0) = \eta(x_1) = 0, \quad (3.3)$$

in (2.1) and to conclude from the vanishing of the first variation of (2.1),

$$\delta J = \varepsilon \frac{\partial J[\bar{y}]}{\partial \varepsilon} \Big|_{\varepsilon=0} = \varepsilon \cdot \int_{x_0}^{x_1} (L_y \eta + L_{y'} \eta') dx = 0, \quad (3.4)$$

(and an integration by parts) that Euler’s equation (3.1) gave a necessary condition for $y(x)$ to render $J[y]$ stationary. (For a detailed and transparent explanation of this vital point see pp. 505–508 of G. F. Simmons, *Differential Equations*, second ed., McGraw-Hill, 1991.) In that paper Lagrange also started working on problems with variable endpoints, with application to brachistochrone and other problems. As another important step forward, he explicitly formulated his *multiplier rule* (without proof). This rule became a basic tool in his *Mécanique analytique*, in which he also included his theory of the calculus of variations and derived from the principle of least action his *equations of motion*, equivalent to *Newton’s second law* and constituting analogues of Euler’s equation. They are:

$$\frac{\partial V}{\partial x_i} + \frac{d}{dt} \left(\frac{\partial T}{\partial x_i} \right) = 0, \quad i = 1, 2, 3, \quad (3.5)$$

where V and T are the potential and kinetic energy, respectively.

Whereas Euler’s interest in the calculus of variations centered around applications, Lagrange’s emphasis was on algorithmic aspects of analysis. Lagrange’s entire work excels by its wealth of original discoveries as well as by the outstanding assimilation of the historical material.

“By generalizing Euler’s method he arrived at his remarkable formulas which in one line contain the solution of all problems of analytical mechanics.

[In his Memoir of 1760–61] he created the whole calculus of variations with one stroke. This is one of the most beautiful articles that have ever been written. The ideas follow one another like lightning with the greatest rapidity . . .”

These enthusiastic lines are taken from lecture notes by C. G. J. Jacobi (1804–51).

4. MINIMAL SURFACES. Euler's *Methodus inveniendi* of 1744 marked not only the beginning of the *theory* of the calculus of variations but also of one of its most fascinating geometric applications related to the creation of a remarkable class of surfaces called *minimal surfaces*. These were originally obtained from the calculus of variations as (portions of) surfaces of least area among all surfaces bounded by a given space curve. Nowadays we define them as surfaces with vanishing mean curvature H ,

$$H = \frac{1}{2}(\kappa_1 + \kappa_2) = \frac{1}{2} \left(\frac{GL - 2FM + EN}{EG - F^2} \right) = 0, \quad (4.1)$$

using the discovery of Meusnier (1756–93) in 1776 that (4.1) is a necessary condition for least area. Here κ_1 and κ_2 are the principal curvatures, and E, F, G and L, M, N are the respective coefficients of the first and second fundamental forms of the surface.

What is most important to us here is Euler's discovery of the first non-trivial minimal surface, the *catenoid*. Euler obtained it by minimizing area, as the surface generated by rotating a *catenary* [a cosh curve, the curve of a hanging chain (*catena*) or cable], say,

$$r = A \cosh x, \quad (4.2)$$

where r is the distance, in 3-dimensional space, from the x -axis.

Euler's discovery of the catenoid was a major accomplishment in his geometric work and marked the beginning of the study of minimal surfaces. It was followed by Lagrange's systematic theory developed in his Memoir of 1760–61. In this paper and in a subsequent one he extended his method to *double integrals* for functions of two variables.

$$J[z] = \int_{\Omega} L(x, y, z, p, q) \, dx \, dy \quad (p = z_x, q = z_y) \quad (4.3)$$

over a domain Ω in the xy -plane subject to given boundary conditions; the corresponding *Euler-Lagrange equation* [taking the place of (3.1)] is

$$L_z - \frac{\partial}{\partial x} L_p - \frac{\partial}{\partial y} L_q = 0. \quad (4.4)$$

5. LEGENDRE, JACOBI, WEIERSTRASS. In the calculus, $y' = 0$ is only a necessary condition for a minimum of a function $y(x)$, and for a decision one must also consider y'' . Similarly, in the calculus of variations Euler's equation is only a necessary condition for a minimum, and for a decision one must also consider the *second variation* of (2.1),

$$\delta^2 J = \frac{\varepsilon^2}{2} \frac{\partial^2 J[\tilde{y}]}{\partial \varepsilon^2} \Big|_{\varepsilon=0} = \frac{\varepsilon^2}{2} \int_{x_0}^{x_1} (L_{yy} \eta^2 + 2L_{yy'} \eta \eta' + L_{y'y'} \eta'^2) \, dx, \quad (5.1)$$

introduced by Legendre (1752–1833) in 1786. It was formally motivated by Taylor's theorem

$$J[y + \varepsilon \eta] = J[y] + \delta J + \delta^2 \tilde{J}, \quad (5.2)$$

where the tilde means that the arguments are $y + \tilde{\varepsilon} \eta$, $y' + \tilde{\varepsilon} \eta'$ with $\tilde{\varepsilon} \in (0, \varepsilon]$. Legendre obtained the condition $L_{y'y'} \geq 0$ along a minimizing curve and $L_{y'y'} \leq 0$ along a maximizing curve (very similar to the calculus!) but he did not justify his analysis completely.

In fact, it took another fifty years before Jacobi succeeded in rigorously demonstrating that $L_{y'y'} > 0$ and the so-called *Jacobi condition*, which asserts that x_1 should be closer to x_0 than the so-called conjugate point* of x_0 , suffice for a local minimum, that is a minimizing \tilde{y} among $y \in C^1[x_0, x_1]$ satisfying the boundary conditions in (2.1) and lying close to \tilde{y} in the C^1 sense, that is, satisfying

$$(a) |y - \tilde{y}| < \rho, \quad (b) |y' - \tilde{y}'| < \rho \quad \text{for small positive } \rho. \quad (5.3)$$

Jacobi's discovery of the conjugate point and of its significance closed a substantial gap. However, condition (5.3b) seemed to be too restrictive and in no way suggested by the nature of the problem. Weierstrass emphasized that one should extend the domain of (2.1) and consider *strong minima*, that is, one should drop (5.3b).

For this program of obtaining a sufficient condition for a strong minimum, Weierstrass set up entirely new machinery centering around two ingenious concepts. The first was a *field of extremals* of (2.1), which he defined as a domain Ω in the xy -plane such that through each of its points there passes precisely one extremal of a one-parameter family of extremals of (2.1) (solution curves of Euler's equation) depending continuously on the parameter, and the second was the so-called *E-function*, a turning point in the history of the calculus of variations.

To define the *E-function*, Weierstrass started from the *slope function* $p = p(x, y)$, the slope at (x, y) of the extremal of a field of extremals $y = h(x, \alpha)$; thus

$$p(x, y) = h'(x, \alpha)|_{\alpha = \alpha(x, y)}, \quad (5.4)$$

where ' refers to differentiation with respect to x . Then he defined the *E-function* by

$$E(x, y, p, y') = L(x, y, y') - L(x, y, p) - (y' - p)L_{y'}(x, y, p), \quad (5.5)$$

where $y = y(x)$ is any C^1 -curve in the region covered by the field of extremals. Now he could prove that if, for an extremal $y = \tilde{y}(x)$ of the field, the above sufficient conditions for a local minimum are satisfied, and if $E \geq 0$ at every point in the field and for every y' , then $\tilde{y}(x)$ gives a strong minimum of (2.1).

In addition to his path-breaking new method, Weierstrass also revolutionized the calculus of variations by stressing—practically for the first time—the importance of a precise definition of the domain $D(J)$ of the functional $J[y]$ and of *admissible functions*, the functions $y \in D(J)$ satisfying the side conditions.

*Department of Mathematics and Statistics
Carleton University
Ottawa, Ontario
Canada K1S 5B6*

*The *conjugate point* of x_0 is the first value $x > x_0$ where a nonzero solution of

$$\frac{d}{dx} \left(L_{y'y'} \frac{dw}{dx} \right) - \left(L_{yy} - \frac{d}{dx} L_{yy'} \right) w = 0, \quad w(x_0) = 0 \quad (x \geq x_0)$$

vanishes. Here $w(x) = \partial y / \partial \alpha|_{\alpha=0}$ and $\alpha = 0$ corresponds to \tilde{y} in the family of extremals $y = y(x, \alpha)$.

THE AUTHORS

DAN KENNEDY received his undergraduate degree at the College of the Holy Cross and his Masters and Ph.D. at the University of North Carolina at Chapel Hill. Since 1973 he has taught mathematics at The Baylor School in Chattanooga, where he holds the Cartter Lupton Distinguished Professorship. He served on the College Board's Advanced Placement Calculus Test Development Committee from 1987 to 1994, the latter four years as chairman, and in June 1994 served as Exam Leader at the AP grading. In the summers he can be found in Algonquin Park, Ontario, where he is program director of a canoe-tripping camp.

JONATHAN L. KING teaches at the University of Florida. His earliest memory of MATHEMATICS—a painful one—was losing a candy bar by failing to find an Eulerian circuit of the spiderweb drawn by a camp counselor. After watching a sullen fellow camper, exhausted by ten seconds of effort, complain “can't *do* it”—and *win* (!) the chocolate from the unscrupulous counselor—he swore to be Prepared for the next Candy-Opportunity, rediscovering Euler's Bridges-of-Königsberg theorem. No candy was proffered for a Bachelor's from Stony Brook, nor for a Ph.D. in 1984 from Stanford in Ergodic Theory with Don Ornstein. An NSF Postdoc provided a stipend—but no chocolate—at College Park and Berkeley. Now an associate professor, he has reached the age where candy is ill-advised.

IGOR RIVIN received his Ph.D from Princeton in 1986 under the supervision of W. P Thurston. His primary research interests are in low-dimensional geometry and topology, but he has also worked in discrete mathematics, mathematical computation and various areas of computer science.

ILAN VARDI received his Ph.D. from MIT in 1982 under the supervision of Dorian Goldfeld. He has held positions at the Institute for Advanced Studies, Stanford University, Wolfram Research Inc., and MSRI. He has also been a visitor at INRIA, Rutgers, and Oklahoma State. His research interests are in Number Theory, Discrete Mathematics, and Computation, all of which were represented in his book “Computational Recreations in *Mathematica*.”

PAUL ZIMMERMANN obtained his Ph.D. thesis in 1991 at École Polytechnique (France). With B. Salvy, he designed a system named Λ -T- Ω (LUO) for the automatic average-case analysis of algorithms. His main research topic is *Algebraic Analysis*, that is, the translation of program specifications into equations of generating functions.

ROGER KRAFT received an Associate of Applied Sciences in Electrical Engineering Technology from Purdue University, a BA in Computer Science and a MA in Mathematics from Indiana University and a Ph.D. in Mathematics from Northwestern University in 1990 under Clark Robinson. Since then he has been a visiting assistant professor at the University of Cincinnati, a visiting assistant professor at Case Western Reserve University, a postdoc at the M.S.R.I., a visiting assistant professor at Pomona College, and is now an assistant professor at Purdue University-Calumet. His field of research is dynamical systems.

JEAN-PAUL ALLOUCHE is at the C.N.R.S. (Centre National de la Recherche Scientifique) in Marseille (France). After having studied mathematics at the École Normale Supérieure de Saint-Cloud, he got his Ph.D. degree at the Université d'Orsay in 1978, and his Doctorat d'État at the Université de Bordeaux in 1983. Prior to his present position, he taught at the I.U.T. d'Informatique in Bordeaux, and the École Normale Supérieure de Fontenay-aux-Roses; later he was employed by the C.N.R.S. in Bordeaux. His research interests are number theory and its relations with theoretical computer science and physics.

DAN ASTOORIAN received his Bachelor of Mathematics, Honours Computer Science (with distinction) from the University of Waterloo in 1992. He is a systems programmer for a software company outside Toronto.

JIM RANDALL graduated from the University of Waterloo with a Bachelor in Mathematics, Honours Computer Science in 1992. He works as a software specialist for Watcom, an Ontario company. In his spare time, he enjoys music and tackling challenging problems. He is currently deciding whether to pursue a graduate degree.

JEFFREY SHALLIT is Associate Professor of Computer Science at the University of Waterloo. He received an A.B. from Princeton University in 1979, and the Ph.D. in Mathematics from the University of California, Berkeley, in 1983. He taught previously at the University of Chicago and Dartmouth College, and was a visiting professor at the Université de Bordeaux and the University of Wisconsin. His book with Eric Bach, *Algorithmic Number Theory*, will be published soon by MIT Press. His interests are in number-theoretic algorithms and the relation between number theory and automata theory.

JEFF ONDICH got his BA in mathematics at St. Olaf College, his Ph.D. in mathematics at the University of Minnesota, and is now masquerading as a computer scientist at Carleton College. Among other things, he is interested in partial differential equations, computational geometry, and computer networks. He likes to hang around with his family, play the piano, watch baseball, and program his 4 kilobyte PDP8/E in binary, but not all at the same time.

MEYER JERISON earned a BS from CCNY, a Master's Degree in applied mathematics from Brown, and a Ph.D. in functional analysis from Michigan. He worked for the predecessor to NASA as well as Lockheed. After a research instructorship at Illinois, he spent forty years at Purdue, retiring in 1991. With Leonard Gillman, he wrote the book *Rings of Continuous Functions*. He has served the MAA as a member of CUPM, the Board of Governors, and the Committee on Publications, and he was book review editor of the Bulletin of the AMS.

Hofstadter's Law: It always takes longer than you expect, even when you take into account Hofstadter's Law.

—Douglas R. Hofstadter
Gödel, Escher, Bach, New York: Basic Books, 1979.

NOTES

(10401) The winding number of a contour C about a point $z_0 \notin C$ is the net number of times C encircles z_0 in the counterclockwise direction. It can be defined using the complex variable $z = x + iy$ as $(2\pi i)^{-1} \int_C dz/(z - z_0)$. The oldest known knight's tour (due to al-Adli in the ninth century) is illustrated below, together with the 49 winding numbers w_{ij} at the corner points between cells. In this example, $\sum w_{ij} = 49$ and $\sum (x_k y_{k+1} - x_{k+1} y_k) = 98$.
 (10402) We use the convention that \mathbb{N} denotes *non-negative* integers. A good reference for products of countably many measure spaces is E. Hewitt and K. Stromberg, *Real and Abstract Analysis*, section 22.

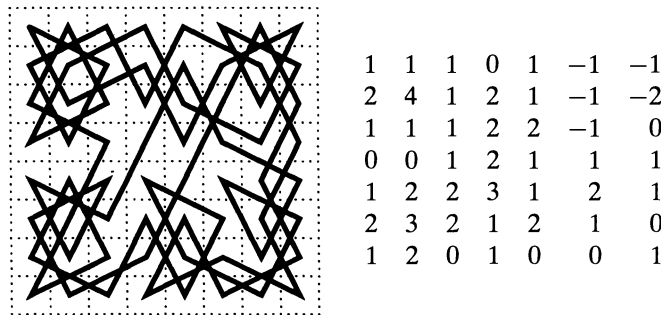


Figure for 10401

SOLUTIONS

Polygonal Polynomial Identities

6639 [1990, 622]. *Proposed by Paul Monsky, Brandeis University, Waltham, MA.*

Let N be a convex centrally symmetric plane n -gon of area 2 divided into n triangles by rays from an interior point P to the vertices. Let a_1, a_2, \dots, a_n be the areas of these triangles, taken in cyclic order.

(i) If N is affinely equivalent to the regular polygon of $2k$ sides, prove that

$$a_1 + a_3 + \dots + a_{2k-1} = a_2 + a_4 + \dots + a_{2k} = 1.$$

(ii) In general prove that a_1, a_2, \dots, a_n satisfy a polynomial identity with integer coefficients and constant term 1.

Solution by John H. Lindsey II, Fort Myers, FL. Evidently $n = 2k$ is even and at least 4. We take all subscripts from the set $[n]$ of integers modulo n . It is convenient to regard N as lying in the xy -plane in cartesian 3-space with the origin O as its center. Let $v_i = (x_i, y_i, 0)$, for $i = 1, 2, \dots, n$ be the position vectors of the vertices of N taken in clockwise order. Since N is centrally symmetric we have

$$v_{i+k} = -v_i \tag{1}$$

for all $i \in [n]$. Let p be the position vector of P and l be the vector $(0, 0, 1)$. Then

$$2a_i = (v_{i+1} - p) \times (v_i - p) \cdot l = v_{i+1} \times v_i \cdot l + (v_i - v_{i+1}) \times p \cdot l \quad (2)$$

for all $i \in [n]$.

In part (i) there is an affine transformation of N onto a regular n -gon N' in the x, y -plane such that N' is also centered on O with area 2. This transformation multiplies all areas by a constant, which must be 1 since N and N' have the same area. Thus the problem is left invariant by this transformation. So we may assume that N is N' . Let S be the set of odd integers modulo n . Then summing either v_i or v_{i+1} over all $i \in S$ is like summing the k^{th} roots of unity in the complex plane, and so yields 0 since $k \geq 2$. Since the $v_{i+1} \times v_i \cdot l$ are all equal to $4/n$, part (i) follows from summing (2) over all $i \in S$.

When $k = 2$, part (ii) follows from part (i), since every convex centrally symmetric plane quadrilateral is affinely equivalent to a square. So in giving the proof of (ii) we may assume $k > 2$.

Summing (2) over $i, i + 1, \dots, i + k - 1$ yields

$$2(a_i + a_{i+1} + \dots + a_{i+k-1}) = (v_{i+1} \times v_i + v_{i+2} \times v_{i+1} + \dots + v_{i+k} \times v_{i+k-1}) \cdot l \\ + (v_i - v_{i+1} + v_{i+1} - v_{i+2} + \dots + v_{i+k-1} - v_{i+k}) \times p \cdot l.$$

Because N is symmetric about O , the first sum on the right hand side of this equality is precisely the area 2 of N . The second sum collapses to $(v_i - v_{i+k}) \times p \cdot l$, which is $2v_i \times p \cdot l$ by (1). Hence

$$v_i \times p \cdot l = a_i + a_{i+1} + \dots + a_{i+k-1} - 1 \quad (3)$$

for all $i \in [n]$.

It follows from (1) and (2) that

$$a_i + a_{i+k} = v_{i+1} \times v_i \cdot l = x_{i+1}y_i - x_iy_{i+1}$$

for all $i \in [n]$. Multiplying this equation by $(-1)^i x_{i+2}x_{i+3} \dots x_{i+k-1}$, summing over $i \in [k]$, and using (1) to give $x_i = -x_{i+k}$, we obtain

$$\sum_{i=1}^k (-1)^i x_{i+2}x_{i+3} \dots x_{i+k-1} (a_i + a_{i+k}) = \\ \sum_{i=1}^k [(-1)^i y_i x_{i+1}x_{i+2} \dots x_{i+k-1} - (-1)^{i+1} y_{i+1}x_{i+2}x_{i+3} \dots x_{i+k-1}x_{i+k}].$$

Evidently this last sum collapses to $-y_1x_2x_3 \dots x_k + (-1)^k y_{k+1}x_{k+2}x_{k+3} \dots x_{2k}$, which is equal to zero by (1). Thus

$$\sum_{i=1}^k (-1)^i x_{i+2}x_{i+3} \dots x_{i+k-1} (a_i + a_{i+k}) = 0. \quad (4)$$

Suppose that $P \neq O$. Rotate the configuration to put P on the positive y -axis. Apply the area-preserving linear transformation $x \mapsto rx, y \mapsto r^{-1}y$ for appropriate r so that p becomes $(0, 1, 0)$. Then (3) gives

$$x_i = a_i + a_{i+1} + \dots + a_{i+k-1} - 1 \quad (5)$$

for any $i \in [n]$. These values of the x_i may be substituted in (4) to obtain an identity satisfied by the a_i . More specifically, let R be the polynomial ring $\mathbb{Z}[A_1, A_2, \dots, A_n]$ in n variables A_1, A_2, \dots, A_n over the integers \mathbb{Z} . For each $i \in [n]$ we define

$$X_i = X_i(A_1, A_2, \dots, A_n) = A_i + A_{i+1} + \dots + A_{i+k-1} - 1 \in R \quad (6)$$

and let e be the polynomial

$$e(A_1, A_2, \dots, A_n) = \sum_{i=1}^k (-1)^i X_{i+2} X_{i+3} \dots X_{i+k-1} (A_i + A_{i+k}) \in R. \quad (7)$$

The above argument tells us that

$$e(a_1, a_2, \dots, a_n) = 0 \quad (8)$$

whenever $P \neq O$.

If $P = O$, then $X_i(a_1, a_2, \dots, a_n) = a_i + a_{i+1} + \dots + a_{i+k-1} - 1 = 0$ for all $i \in [n]$, since N is symmetric about O with area 2. In view of (7) this implies that (8) holds when $P = O$. Thus (8) holds in all cases. Of course, the fact that the area of N is 2 says that

$$f(a_1, a_2, \dots, a_n) = 0, \quad (9)$$

where f is the polynomial

$$f(A_1, A_2, \dots, A_n) = A_1 + A_2 + \dots + A_n - 2. \quad (10)$$

It follows from (6) that $X_i + X_{i+1} \equiv A_i + A_{i+k} \pmod{2R}$ and $X_i + X_{i+k} = f$ for any $i \in [n]$. We conclude that $X_i \equiv -X_{i+k} \equiv X_{i+k} \pmod{2R + fR}$ for all such i , and hence that

$$\begin{aligned} e &\equiv \sum_{i=1}^k X_{i+2} X_{i+3} \dots X_{i+k-1} (X_i + X_{i+1}) \\ &\equiv \sum_{i=1}^k X_{i+1} X_{i+2} \dots X_{i+k-1} + \sum_{i=1}^k X_{i+2} X_{i+3} \dots X_{i+k} \\ &\equiv X_2 X_3 \dots X_k + X_{k+2} X_{k+3} \dots X_{2k} \equiv 0, \end{aligned}$$

where all equivalences are modulo the ideal $2R + fR$ of R . Thus there is some polynomial $g \in R$ such that $h = e - fg$ has all coefficients even.

Let $e(0)$, $f(0)$, $g(0)$ and $h(0)$ be the constant terms of e , f , g and h , respectively. Since $f \equiv A_1 + A_2 + \dots + A_n \pmod{2R}$, the terms of degree one in fg are congruent to $g(0)(A_1 + A_2 + \dots + A_n)$ modulo $2R$. The terms of degree one in e are

$$\sum_{i=1}^k (-1)^{i+k} (A_i + A_{i+k}) \equiv A_1 + A_2 + \dots + A_n \pmod{2R},$$

while those of h are all even. Since $e = fg + h$, we conclude that $g(0)$ is odd. It follows that $h(0) = e(0) - g(0)f(0) = 2g(0)$ is exactly divisible by 2. Thus $h/2 - jf \in R$ has constant term 1 for some $j \in \mathbb{Z}$. Because e , f and h all vanish at (a_1, a_2, \dots, a_n) , the polynomial $h/2 - jf$ provides a solution to part (ii) of the problem.

For example, in the case $k = 3$ we get

$$e = (1 - A_3 - A_4 - A_5)(A_1 + A_4) + (A_4 + A_5 + A_6 - 1)(A_2 + A_5) + (1 - A_5 - A_6 - A_1)(A_3 + A_6)$$

and we may take $g = 1 + A_4 + A_5 + A_6$ and $j = 0$. This gives the polynomial

$$\begin{aligned} \frac{1}{2}h - jf &= \frac{1}{2}(e - fg) \\ &= 1 - A_2 + A_4 + A_6 - A_4^2 - A_6^2 \\ &\quad - A_1(A_3 + A_4 + A_5 + A_6) \\ &\quad - A_3(A_4 + A_5 + A_6) - A_4(A_5 + A_6) - A_5A_6, \end{aligned}$$

which provides a solution of part (ii) for $k = 3$.

Solved also by the proposer. Part (i) only was solved by Weiqi Gao.

The Difference between Graphs of Even and Odd Size

6673 [1991, 965]. *Proposed by Paresh J. Malde and Allen J. Schwenk, Western Michigan University, Kalamazoo, MI.*

Let $c_n = e_n - d_n$, where e_n is the number of connected labeled n -vertex graphs with an even number of edges and d_n is the number of connected labeled n -vertex graphs with an odd number of edges. For example, $c_2 = 0 - 1 = -1$ and $c_3 = 3 - 1 = 2$. Find a general formula for c_n .

Composite solution I by Stephen C. Locke, Florida Atlantic University, Boca Raton, FL, and Herbert S. Wilf, University of Pennsylvania, Philadelphia, PA. The answer is $c_n = (-1)^{n-1}(n-1)!$. More generally, for a fixed positive integer k , let $c_{n,k}$ be the same excess, but for graphs with exactly k components. We claim that $c_{n,k}$ is the signed Stirling number of the first kind.

Let $f(n, e, k)$ be the number of labeled graphs with n vertices, e edges, and k components, and form the exponential generating functions

$$F_k(x, y) = \sum_{n,e} f(n, e, k) \frac{x^n y^e}{n!}.$$

The computation also uses the corresponding generating function for all labeled graphs,

$$F(x, y) = \sum_{n,e} f(n, e) y^e x^n / n!,$$

where $f(n, e)$ is the number of labeled graphs with n vertices and e edges, so that $f(n, e) = \binom{C(n,2)}{e}$ with $C(n, 2) = \binom{n}{2}$. Then by the Exponential Formula we have

$$F_k(x, y) = \frac{F_1(x, y)^k}{k!} \text{ and } F(x, y) = \exp(F_1(x, y))$$

as formal series. For fixed k , the exponential generating function of $c_{n,k}$ is $F_k(x, -1)$.

The formula for $f(n, e)$ yields $F(x, y) = \sum_{n \geq 0} (1+y)^{C(n,2)} x^n / n!$, so that $F(x, -1) = 1 + x$. Thus, $F_1(x, -1) = \ln(1+x)$ and $F_k(x, -1) = \ln(1+x)^k / k!$. These are known exponential generating functions of the claimed values.

Composite solution II by Bruce E. Sagan, Michigan State University, East Lansing, MI, and Richard Holzsager, The American University, Washington, DC. We construct a bijective proof of the formula $c_n = (-1)^{n-1}(n-1)!$ by associating graphs in pairs that are identical everywhere except on the pair $\{v_1, v_2\}$; one has this pair as an edge and the other does not. Note that every n -vertex graph belongs to such a pair. If both graphs in a pair are connected or both are disconnected, then the pair contributes zero to c_n .

In computing c_n , we may therefore restrict our attention to connected graphs containing the edge $v_1 v_2$ as a cut-edge. Let $z = v_1$. Since $z v_2$ is a cut-edge, z and v_2 have no common neighbor. Hence contracting v_2 into z by contracting the edge $z v_2$ reduces the number of edges in every such graph by exactly one.

Among these graphs, we consider those that do or do not contain the edge $z v_3$. All are paired off with zero contribution to c_n except those in which $z v_3$ is a cut-edge. Again we can contract this cut-edge in the remaining graphs, absorbing v_3 into z . We continue this process until the graphs that remain have been contracted to a single vertex.

We claim a graph survives the pairing at each stage if and only if for each $j > 1$ the vertex v_j has exactly one edge to a vertex with lower index. If some v_j has two lower neighbors, then at the moment when the higher of these would be absorbed by contraction, it would have a common neighbor with z . If some v_j has no lower neighbor, let P be a shortest path

from v_j to a vertex with lower index; any vertex along P whose index is a local maximum now has two lower neighbors.

The number of graphs satisfying this structural property is $(n - 1)!$, and they all have $n - 1$ edges, so the desired formula holds.

Editorial comment. Use of the generating function, as in Solution I, is the more straightforward approach. Indeed, these generating functions are well known in the literature of graph theory. The Exponential Formula may be found in H. S. Wilf, *Generatingfunctionology*, Academic Press, 1990. Also see I. P. Goulden and D. M. Jackson, *Combinatorial Enumeration*, Wiley, 1983 (where the problem of finding the relation between the generating functions $F_1(x, y)$ and $F(x, y)$ is given as Problem 3.4.7 on p. 210), and the review paper, E. M. Palmer, "The enumeration of graphs", pp. 385–415 in L. W. Beineke and R. J. Wilson, eds., *Selected Topics in Graph Theory*, Academic Press, 1978 (where this result may be found on p. 391). In all, eight solutions to the present problem of this type were submitted.

V. A. Liskovets notes that similar results hold for different classes of graphs, such as digraphs allowing edges in both directions, digraphs not allowing such edges, graphs with loops, and even strongly connected digraphs. He supplied a reference to V. A. Liskovets, *Doklady AN BSSR* 17 (1973), 1079.

Bijjective proofs, as in Solution II, were submitted by four solvers.

Solved also by G. Calinescu (student, Romania), R. J. Chapman (U. K.), F. Galvin, V. A. Liskovets (Belarus), R. Martin (student), J. Mobek, F. Schmidt, and the proposers.

Squarish Numbers

10220 [1992, 461]. *Proposed by Solomon W. Golomb, University of Southern California, Los Angeles, CA.*

Suppose ϵ is a given positive number. A positive integer n will be called ϵ -suarish if and only if it has a factorization $n = ab$ with $1 \leq a < b < (1 + \epsilon)a$. Prove that there are infinitely many occurrences of six consecutive ϵ -suarish numbers.

Solution by Robert B. Israel, University of British Columbia, Vancouver, B.C., Canada. If x is sufficiently large, then $(x-1)(x+1) = x^2 - 1$ and $(x-2)(x+2) = x^2 - 4$ are both ϵ -suarish, as is x^2 itself. Also, if $x = t^2 + t - 2$ for some integer t , then $(x-t)(x+t+1) = x^2 - 2$ and $(x-2t+1)(x+2t+3) = x^2 - 5$ are both ϵ -suarish for t sufficiently large. Finally, if $x = 2s^2 - 2$ for some integer s , then $(x-2s+1)(x+2s+1) = x^2 - 3$ is ϵ -suarish for s sufficiently large.

This gives 6 consecutive ϵ -suarish integers $x^2 - 5, x^2 - 4, \dots, x^2$ if $x = t^2 + t - 2 = 2s^2 - 2$ with s and t sufficiently large. It suffices to show that $u^2 = 8s^2 + 1$ has solutions with arbitrarily large s and $u = 2t + 1$. One could use the theory of Pell's equation here or simply take

$$\begin{pmatrix} s \\ u \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 8 & 3 \end{pmatrix}^n \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

for large positive integer n .

Editorial comment. All solutions used essentially the same construction, and none of them considered longer sequences of consecutive ϵ -suarish numbers. To illustrate the construction, both Richard Holzsager and the proposer mentioned the numerical example

$$55 \cdot 89 (= 4895), \quad 68 \cdot 72, \quad 59 \cdot 83, \quad 62 \cdot 79, \quad 69 \cdot 71, \quad 70 \cdot 70 (= 4900)$$

for $\epsilon > 34/55$. The proposer declined to conjecture arbitrarily long runs, but does not expect that this problem gives the best possible result.

Solved also by C. P. Grant, R. Holzsager, I. Kastanas, and the proposer.

The Perimeter of an Oval

10227 [1992, 463]. *Proposed by Antonio Montes, Universitat Politècnica de Catalunya, Barcelona, Spain.*

Suppose that α is a convex simple closed curve in the plane which is piecewise C^1 and suppose that the origin lies inside α .

(a) Show that the length of α is given by

$$-\oint_{\alpha} \vec{\rho} \cdot d\vec{\tau},$$

where $\vec{\tau}$ is the unit tangent vector in a counterclockwise direction and $\vec{\rho}$ is the vector from the origin to the curve.

(b) If, in addition, α is C^1 and piecewise C^2 , show that the length of α is given by

$$\int_0^{2\pi} r(\theta)^2 \kappa(\theta) d\theta,$$

where $r(\theta)$ indicates the distance from the origin to the point on α with polar angle θ and $\kappa(\theta)$ denotes the curvature of α at the point $(\theta, r(\theta))$.

Solution I by Richard Holzsager, The American University, Washington, DC. (a) One representation of the standard integral for length is $\int d\vec{\rho} \cdot \vec{\tau}$. Suppose α is broken by points $\rho_0, \rho_1, \dots, \rho_n = \rho_0$ into C^1 segments $\alpha_1, \alpha_2, \dots, \alpha_n$. Integrating by parts on the segment α_j gives

$$\text{length}_j = \int_j d(\vec{\rho} \cdot \vec{\tau}) - \int_j \vec{\rho} \cdot d\vec{\tau} = \vec{\rho}_j \cdot \vec{\tau}_{j,1} - \vec{\rho}_{j-1} \cdot \vec{\tau}_{j,0} - \int_j \vec{\rho} \cdot d\vec{\tau}$$

where the subscripts 0 and 1 on τ refer to the values at the two ends of the segment. If we sum over j , and let $d\vec{\tau}_j$ represent the difference between the outgoing tangent $\vec{\tau}_{j+1,0}$ and the incoming tangent $\vec{\tau}_{j,1}$, then summation gives

$$\text{length} = -\sum \int_j \vec{\rho} \cdot d\vec{\tau} - \sum \vec{\rho}_j \cdot d\vec{\tau}_j = -\int \vec{\rho} \cdot d\vec{\tau},$$

provided we interpret the last integral in the Riemann-Stieltjes sense, taking the gaps $d\vec{\tau}_j$ into account.

(b) Let $\vec{\rho}(\theta) = r(\theta)e^{i\theta}$ in complex notation. Since the curve is convex and surrounds the origin the normal \vec{v} equals $i\vec{\tau}$.

Hence

$$\vec{\tau}(s) = \frac{d\vec{\rho}}{d\theta} \frac{d\theta}{ds} = \frac{d\theta}{ds} (r'(\theta)e^{i\theta} + ir(\theta)e^{i\theta})$$

and

$$\vec{v}(s) = i\vec{\tau}(s) = \frac{d\theta}{ds} (ir'(\theta)e^{i\theta} - r(\theta)e^{i\theta})$$

where s is the arclength parameter. But $\vec{\tau}'(s) = k(s)\vec{v}(s)$ which implies

$$\vec{\rho}(s) \cdot \frac{d\vec{\tau}}{ds} = k(s)\vec{\rho}(s) \cdot \vec{v}(s) = -\frac{d\theta}{ds} r(\theta(s))^2 k(\theta(s))$$

and (b) then follows from (a) by the transformation of integral formula.

Solution II by G. D. Chakerian, University of California, Davis, CA, and W. J. Firey, Oregon State University, Corvallis, OR. Let K denote a plane convex body with interior points; write α for the boundary of K . Choose rectangular coordinates (x, y) such that $(0, 0)$ is an interior point of K . Assign the counter-clockwise orientation to α .

For each ϕ , $0 \leq \phi < 2\pi$ there is a unique half-plane of points (x, y) such that

$$x \cos \phi + y \sin \phi \leq h(\phi) \quad (1)$$

contains K and its bounding line $L(\phi)$ meets α but not the interior of K . $L(\phi)$ is the support line to K which corresponds to the outer normal direction $\vec{v}(\phi) = (\cos \phi, \sin \phi)$ to (1) and to K and α ; $h(\phi)$ is the positive distance from $(0, 0)$ to $L(\phi)$. Orient $L(\phi)$ consistently with the orientation of α . Thus $\vec{\tau}(\phi) = (-\sin \phi, \cos \phi)$ is a generalized forward tangent direction corresponding to $\vec{v}(\phi)$.

$K \cap L(\phi)$ is either a point $(x(\phi), y(\phi))$ or a closed segment of points

$$(x(\phi) - \lambda \sin \phi, y(\phi) + \lambda \cos \phi), \quad 0 \leq \lambda \leq l,$$

where l is the length of $K \cap L(\phi)$. Define $\vec{\rho}(\phi) = (x(\phi), y(\phi))$, that is either the point $K \cap L(\phi)$, or the initial point of $K \cap L(\phi)$. Note that

$$(x(\phi) - \lambda \sin \phi) \cos \phi + (y(\phi) + \lambda \cos \phi) \sin \phi = \vec{\rho}(\phi) \cdot \vec{v}(\phi) = h(\phi) \quad (2)$$

for all allowed λ . This shows that, when $K \cap L(\phi)$ is a segment, any systematic definition of $\vec{\rho}(\phi)$ as a point on that segment will produce the same result. Also

$$d\vec{\tau}(\phi) = (-\cos \phi, -\sin \phi)d\phi = -\vec{v}(\phi)d\phi. \quad (3)$$

The convexity of K ensures that all discontinuities of $\vec{\rho}(\phi)$ are finite jumps of the form

$$\lim_{\phi \rightarrow \phi_0+} (\vec{\rho}(\phi) - \vec{\rho}(\phi_0)) = (x(\phi_0) - l \sin \phi_0, y(\phi_0) + l \cos \phi_0) - (x(\phi_0), y(\phi_0)) = l \vec{\tau}(\phi_0)$$

at ϕ_0 . Each corresponds to a segment of length l in α ; the set of these discontinuities is countable and the corresponding sum of length l is finite. This is all due to the convexity of K . Hence any product $\vec{\rho}(\phi) \cdot \vec{\mu}(\phi)$, $\mu(\phi)$ continuous, is integrable over $0 \leq \phi < 2\pi$.

Part (a) of the problem requires us to prove that the length of α is

$$- \int_{\alpha} \vec{\rho} \cdot d\vec{\tau}. \quad (4)$$

Here α , $\vec{\rho}(\phi)$, $\vec{\tau}(\phi)$ have the meanings given above. Any one of the ways of defining the length of α (e.g. by Hausdorff measure, integral-geometric procedures, or as a limit of lengths of approximating polygons) leads to the representation of the length of α as

$$\int_0^{2\pi} h(\phi) d\phi \quad (5)$$

valid for any α , since the convexity of K entails the continuity of $h(\phi)$. Thus it suffices to prove that (4) equals (5). This is immediate from (2) and (3). Note that the restrictions in the statement of part (a) of the problem are not needed.

Part (b) is obtained from part (a) by transforming the integral as in Solution I. Here the assumption that α is piecewise C^2 is used to provide integrable expressions in the formulas.

Editorial comment. Solution II shows that one can obtain a stronger result by exploiting the convexity. This approach was also taken by H. W. Guggenheimer. The other solvers followed the approach of Solution I.

Solved also by J. Anglesio (France), R. J. Chapman (U. K.), H. W. Guggenheimer, A. Nijenhuis, and the proposer.

10261 [1992, 872]. *Proposed by Wu Wei Chao, He Nan Normal University, Xin Xiang City, He Nan Province, China.*

Let x be a real number such that $0 < x < \pi/4$. Prove that

$$(\sin x)^{\sin x} < (\cos x)^{\cos x}.$$

Solution I by Alain Salinier, Université de Limoges, Limoges, France. We produce a positive lower bound on $\ln(\cos x) - \tan x \ln(\sin x)$ for $0 < x < \pi/4$. For all real numbers a, b , and λ with $a > 0, b > 0$ and $\lambda \in (0, 1)$, the strict concavity of the natural logarithm gives

$$\ln(\lambda a + (1 - \lambda)b) > \lambda \ln a + (1 - \lambda) \ln b. \quad (1)$$

Let $x \in \mathbb{R}$ with $0 < x < \pi/4$ and define $a = \sin x, b = \sin x + \cos x, \lambda = \tan x$. The above conditions are satisfied, so (1) gives

$$\ln(\cos x) > \tan x \ln(\sin x) + (1 - \tan x) \ln(\sin x + \cos x).$$

The last term is positive since $\tan x < 1$ and $\sin x + \cos x = \sqrt{2} \cos(\pi/4 - x) > 1$.

Solution II by Ilias Kastanas, California State University, Los Angeles, CA. Let $u = \sin^2 x$ and $v = \cos^2 x$. Then $u + v = 1$ and $0 < u < 1/2 < v < 1$. The desired inequality is equivalent to the following chain of inequalities:

$$\begin{aligned} \sqrt{u} \ln u &< \sqrt{v} \ln v \\ \sqrt{u} \left(v + \frac{v^2}{2} + \frac{v^3}{3} + \dots \right) &> \sqrt{v} \left(u + \frac{u^2}{2} + \frac{u^3}{3} + \dots \right) \\ \sqrt{v} \left(1 + \frac{v}{2} + \frac{v^2}{3} + \dots \right) &> \sqrt{u} \left(1 + \frac{u}{2} + \frac{u^2}{3} + \dots \right). \end{aligned}$$

Since all coefficients and exponents are positive, the last series represents an increasing function, and the result follows.

Solution III by Heinz-Jürgen Seiffert, Berlin, Germany. Since $0 < \tan x < 1$, the weighted arithmetic-geometric mean inequality gives

$$\left(1 + \tan^2 x\right)^{1 - \tan x} \left(\tan^2 x\right)^{\tan x} < (1 - \tan x)(1 + \tan^2 x) + \tan x \tan^2 x. \quad (2)$$

The left side of (2) is $(\sin^2 x)^{\tan x} / \cos^2 x$, and the right side is $1 + \tan^2 x - \tan x < 1$. Hence $(\sin^2 x)^{\tan x} < \cos^2 x$, from which the desired result follows by raising both sides to the power $(\cos x)/2$.

Solution IV by Mark A. Pinsky, Northwestern University, Evanston, IL. Write $a = \sin x$ and $b = \cos x$. We must show $a \ln a < b \ln b$ for $0 < x < \pi/4 \approx .7853981635$. The function $f(u) = u \ln u$ is convex for $0 \leq u$ with a minimum at $u = e^{-1}$ and $f(e^{-1}) = -e^{-1}$. If $x \geq \arcsin(e^{-1}) \approx .3767275081$, then $e^{-1} \leq a < b$ implies $f(a) < f(b)$ as required. Otherwise, $(a, f(a))$ is below the chord joining $(0, 0)$ to $(e^{-1}, -e^{-1})$, and $(b, f(b))$ is above the tangent to $y = f(x)$ at $(1, 0)$. Thus $a \ln a < -a, b \ln b > b - 1$, and $b \ln b - a \ln a > b - 1 + a$. Since $(a + b)^2 = 1 + 2ab > 1$, the result follows.

Note that we have shown the more general result that $0 < a < b$ and $a + b > 1$ implies $a^a < b^b$.

Solved by 81 readers (including those cited) and the proposer. Two incorrect solutions were also received.

Collaborating editors: David F. Appleyard, Paul T. Bateman, Bruce C. Berndt, Duane M. Broline, Barry W. Brunson, Frank S. Cater, Gulbank D. Chakerian, Underwood Dudley, Gerald A. Edgar, Michael A. Filaseta, Ira M. Gessel, Richard A. Gibbs, Jerrold R. Griggs, Douglas A. Hensley, John R. Isbell, Mourad E. H. Ismail, Murray Klamkin, Daniel J. Kleitman, Frederick W. Luttman, Frank B. Miles, Richard Pfiefer, Stephen L. Portnoy, J. O. Shallit, John Henry Steelman, Kenneth B. Stolarsky, David E. Tepper, Douglas B. Tyler, Daniel Ullman, and William E. Watkins.

Two EZ Proofs of $\sin^2 z + \cos^2 z = 1$

Leon EHRENPREIS and Doron ZEILBERGER

The two proofs contrast two origins of the sine function: (a) the zeros of $\sin z$, and (b) the power series for $\sin z$.

For the proof based on (a), we study the zeros of $f(z) := \sin^2 z + \cos^2 z - 1$. Observe that f vanishes when $z = n\pi$ or $(n + \frac{1}{2})\pi$, for any integer n . This is the Pythagorean theorem for degenerate triangles. But, even more, f has a second order zero at these points, because f is even (so it must have a double zero at $z = 0$), and periodic of period $\pi/2$.

Now we appeal to a theorem on entire functions. $f(z)$ is an entire function of exponential type 2. In fact $|f(z)| \leq Ce^{2|z|}$. It is a standard consequence of the argument principle that such a function cannot have more than $(2 + \epsilon)r/\pi$ zeros (counting multiplicities) in $|z| \leq r$, unless it vanishes identically. But we have produced $4r/\pi$ zeros there. Thus $f(z) \equiv 0$. \square

The same method of proof can be used to prove many identities for elliptic functions.

For the proof based on (b), observe that $\sin z$ (resp. $\cos z$) is the *exponential generating function* (henceforth e.g.f.)¹ for increasing sequences of integers of odd (resp. even) size, weighted by $(-1)^{\lfloor \text{size}/2 \rfloor}$. Hence $\sin^2 z + \cos^2 z$ is the e.g.f. for ordered pairs of increasing sequences of integers $(a_1 < \dots < a_r; b_1 < \dots < b_s)$ such that $r + s$ is even, $\{a_1, \dots, a_r, b_1, \dots, b_s\} = \{1, 2, \dots, r + s\}$, and the weight is $(-1)^{\lfloor r/2 \rfloor + \lfloor s/2 \rfloor}$. Let the *mate* of such a pair be $(a_1, \dots, a_r, b_s; b_1, \dots, b_{s-1})$ if $a_r < b_s$, and $(a_1, \dots, a_{r-1}; b_1, \dots, b_s, a_r)$ if $a_r > b_s$. Since every pair has opposite sign from its mate, the total weight of each couple is 0. The only left-over is the pair (*empty, empty*) that has no mate; its weight is 1, and its size is 0, hence its e.g.f. is 1. \square

A similar (but not identical) proof was found independently by Ed Scheinerman and will appear in *Math Magazine*.

Temple University

Philadelphia, PA 19122

[leon,zeilberg]@math.temple.edu.

Nisan 12, 5754.

¹ The e.g.f. of a combinatorial family of labelled objects according to a weight w is $\sum_{n=0}^{\infty} a_n z^n / n!$, where a_n is the sum of the weights of all the objects of size n . The product $A \times B$ of two such combinatorial families is the set of ordered pairs (a, b) , $a \in A$, $b \in B$, with the labels of a and b disjoint, $\text{size}(a, b) := \text{size}(a) + \text{size}(b)$, and $\text{weight}(a, b) = \text{weight}(a) \text{weight}(b)$. It is easy to see that the e.g.f. of $A \times B$ is the products of the e.g.f.s of A and B . See D. Foata, M. Schutzenberger, LNM # 138 (Springer), and H. Wilf, "generatingfunctionology", Academic Press.

REVIEWS

Edited by **Darrell Haile**
Indiana University, Bloomington IN 47405

<p><i>How to Teach Mathematics.</i> By Steven G. Krantz. American Mathematical Society, 1993, v + 76 pp., \$21.00.</p>
--

Reviewed by Meyer Jerison

Addressing teachers of college mathematics, Steven G. Krantz proposes in *How to Teach Mathematics* "to set down the traditional principles of good teaching in mathematics" as he sees them. "While perhaps most experienced mathematics instructors would agree with much of what is in this booklet," he says, "in the final analysis this tract must be viewed as a personal polemic on how to teach."

His central message is that good teaching is important not only for the administration and the students but also for the morale of the teacher and that thought and effort devoted to preparation will pay off handsomely in time, effort, and satisfaction in performance. In 47 short sections, he makes concrete suggestions, aimed primarily at the inexperienced instructor, on how to be a good teacher. He encourages instructors to choose methods for themselves: "If you do not agree in every detail with what I say, then I hope that at least my remarks will give you pause for thought. In the end, you must decide for yourself what will take place in your classroom."

First and above all he emphasizes instructors' respect for themselves as well as for their audience, noting that "exercising patience requires no more effort than exercising your vocal chords with an insulting remark." He predicts that "things will go smoothly if the attitude in your class is that you and the students are working together to conquer the material ... make it clear from the outset that you are on the student's side." Again and again he urges the instructor to be thoroughly prepared and to speak and write clearly. He strongly favors encouraging questions from students and answering them with respect, and he gives examples of good and bad ways to answer questions.

As an example of his ideas on presentation I cite his criticism of the formulation in many calculus books of Green's theorem with special hypotheses on the domain. "This is because the authors are looking ahead to the proof, and want to state the theorem in precisely the form in which it will be proved. The entire approach is silly. Why not state Green's theorem in complete generality? Then it is simple, sweet, and students can see what the principal idea is. When it is time for the proof, just say 'to keep the proof simple, and to avoid technical details, we restrict attention to a special class of domains'."

He is a skeptic on innovations in teaching. "This author, and this booklet," he says, "has a built in bias toward traditional methods, such as lectures ... There are those who will criticize this book for being reactionary. I welcome their remarks." He is clearly not an admirer of professional education. He does, nevertheless,

endorse the periodical *UME Trends*, saying "It is important that we explore new methods to teach mathematics."

He has evidently thought seriously about computers and programmable calculators in mathematics courses and claims that we do not yet know how they can best contribute to learning. He is convinced that they are being oversold, and asserts: "Using the quadratic formula is easy. Analyzing word problems is hard. A person who cannot do the first will also probably not be able to do the second—with or without the aid of a machine."

I think that virtually everyone will agree with what has been reported so far as the author's advice or will grant him the right to his personal opinion. Some helpful advice might be added to his.

In preparation for a lecture, for example, an inexperienced teacher should be warned that choice of notation can be critical. It is altogether too easy, when doing a problem, to get stuck in a corner with inconsistent notation. The students should be told how and why some of the choices are made.

Krantz objects vigorously and properly to an examination where the average score is 32%. He says that students lack the perspective to estimate their standing, for example, by doubling their scores. But I would go even further and say that no reasonable student will equate a grade of 45% with one of 90%. The message of a grade of 45% is clearly that there is a huge gap in the student's grasp of the material, and no amount of rationalization can change that.

To the section entitled *Why Do We Need Mathematics Teachers?* I would add two reasons. One is that the teacher provides a pace for the student. Most people need help in setting a reasonable pace for themselves when they try to master difficult material. The second is that the teacher aims to improve the student's ability to read. The notion that some students cannot read while others can read obscures the reality that everybody (professors included) can read some material and cannot read other material. If I were to set a single criterion for the aim of a college education it would be to *expand* every student's ability to read. Ideally, education does not end with college. The broader the range of material that a person can read by the time of graduation from college the more likely that that education will continue for the rest of her/his life.

I do have some serious objections to the book. A major one is that the author does not apply his own "principles of good teaching" to the book itself. "The first few lectures . . . set the tone for the entire semester" he says, and "respect . . . the audience." But the longest paragraph on the very first page of the preface is devoted to a caricature of the unenthusiastic professor who gives "a lecture ranging from dreary to arrogant to boring to calamitous." That paragraph made me uncomfortable just as I was beginning to read the book, and I shuddered at the thought of seeing it quoted (partially) in some publication that trashes college teaching and attributes it to the publisher, the prestigious American Mathematical Society.

In a similar vein, quite properly he warns against sarcasm or saying anything that will diminish a student's self respect. But he says to the reader: "If that is not your role as teacher then what, pray tell, could it possibly be?" He warns: "sometimes the professor just doesn't realize that he/she is behaving in a manner that some students take amiss," but there are statements in the booklet that some readers will take amiss. His remarks about education departments and about math anxiety are examples.

The publisher likes the following quote from the preface so much that it is repeated on the outside cover: "The good news is that it requires no more effort,

no more preparation, and no more time to be a good teacher than to be a bad teacher. The proof is in this booklet.” However, the author is realistic enough to contradict it often. On page 2 he says: “Throughout this booklet, I will repeatedly exhort you to prepare your lectures.” An entire section is entitled *Prepare*. It begins with a quotation from the bad teacher: “My time is too valuable. I am not going to spend it preparing.” Then the author says: “I cannot over-emphasize the fact that preparation is of utmost importance if you are going to deliver an effective lecture or give a stimulating class.” (He also warns against over-preparation.) In the section *Lectures*, he says: “you really have to work at making your lectures reach your students.” He then criticizes the tendency “not to put a lot of effort into our teaching.” I believe that a case can be made for the proposition that preparation, time, and effort will pay off in the long run, and the good habits that the author urged teachers to develop will eventually lead to more effective teaching with less effort and more satisfaction than the minimal attention to teaching that he criticizes. But the booklet is “primarily for the graduate student or novice instructor” who is being urged to devote more effort, more preparation, and more time than is needed just to get by in the classroom.

As the subtitle of the booklet suggests, the author does not refrain from talking about himself, and that imparts a liveliness to the text, but it also has some curious consequences. At the beginning of the preface he writes of deterioration with age of the quality of teachers and says: “and I speak here of myself as much as anyone.” But three pages later he writes: “I find that my teaching gets better and better.”

In spite of the reservations that I have expressed, I can recommend that beginners read this book. They will find many useful suggestions for improving their effectiveness and consequently their comfort in the classroom. They will also be encouraged to think seriously about what they are doing.

*Department of Mathematics
Purdue University
West Lafayette, IN 47907-1395*

One of the big misapprehensions about mathematics that we perpetrate in our classrooms is that the teacher always seems to know the answer to any problem that is discussed. This gives students the idea that there is a book somewhere with all the right answers to all of the interesting questions, and that teachers know those answers. And if one could get hold of the book, one would have everything settled. That's so unlike the true nature of mathematics.

—Leon Henkin

Teaching Teachers, Teaching Students, by L. A. Steen and D. J. Albers, eds., Boston: Birkhauser, 1981, p. 89.

TELEGRAPHIC REVIEWS

Edited by **Arnold Ostebee and Paul Zorn**

with the assistance of the Mathematics Departments of
Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T : Textbook	P : Professional Reading	1-4 : Semester
C : Computer Software	L : Undergraduate Library	** : Special Emphasis
S : Supplementary Reading	13 : Grade Level	?? : Questionable

Readers are advised that price information is subject to change. Selected books and software packages receive a second, more extensive review in the *Monthly*.

Books and software submitted for review should be sent to *Book Reviews Editor, American Mathematical Monthly, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098.*

General, S(15-16). *Der Goldene Schnitt.* Hans Walser. BG Teubner Leipzig, 1993, 140 pp, DM 16,80 (P). [ISBN 3-8154-2070-9] The golden mean as it appears in geometry and other areas (e.g., fractals, number sequences, probability). JD-B

General, S*(13-16), L.** *Newton's Clock: Chaos in the Solar System.* Ivars Peterson. WH Freeman, 1993, xiii + 317 pp, \$22.95. [ISBN 0-7167-2396-4] An inviting history of celestial mechanics from antiquity (e.g., navigational star charts) to present (e.g., chaos vs. stability in the solar system). Informal but informative, learned but readable. PZ

General, S(15). *Mathematik für Wirtschaftswissenschaftler.* Volker Nollau. BG Teubner Leipzig, 1993, 260 pp, DM 29,80 (P). [ISBN 3-8154-2046-6] An introduction to mathematics germane economics. Set theory, propositional calculus, number systems, linear algebra and optimization, sequences and series, calculus, linear difference and differential equations, probability, random variables, probability distributions. JD-B

General, T(13-16), S*, P, L**.** *The USSR Olympiad Problem Book: Selected Problems and Theorems of Elementary Mathematics.* D.O. Shklarsky, N.N. Chentzov, I.M. Yaglom. Transl: John Maykovich. Dover, 1993, xvi + 452 pp, \$10.95 (P). [ISBN 0-486-27709-7] Unabridged republication of the 1962 W.H. Freeman edition—among best problem collections ever. 320 challenging problems with unusually complete and instructive solutions. A real bargain. LCL

Finite Mathematics, T(13: 1). *Finite Mathematics, Second Edition.* Stanley I. Grossman. Wm C Brown, 1993, xviii + 685 pp, \$60.63. [ISBN 0-697-11351-5] Standard topics: sets, linear algebra, linear programming, counting, probability, statistics, Markov chains, finance. Many carefully worked examples; applications to business, biology. Uses real data. (1983 Wadsworth text, TR, October 1983.) TH

Education, P. *Advances in Instructional Psychology, Volume 4.* Ed: Robert Glaser. Lawrence Erlbaum Assoc, 1993, xi + 347 pp, \$79.95. [ISBN 0-8058-0709-8] Reports on research on reasoning and problem solving as fundamental components of learning and teaching. Two chapters focus on mathematics. MW

Education, P. *Handbook of Individual Differences, Learning, and Instruction.* David H. Jonassen, Barbara L. Grabowski. Lawrence Erlbaum Assoc, 1993, xvii + 488 pp, \$34.50 (P); \$99.95. [ISBN 0-8058-1413-2; 0-8058-1412-4]

Education, P. *Schools for Thought: A Science of Learning in the Classroom.* John T. Bruer. MIT Pr, 1993, x + 324 pp, \$29.95. [ISBN 0-262-02352-0] Clear introduction to cognitive science, with persuasive arguments for its importance in educational reform. Summaries of representative research programs, examples of specific applications, implications for teacher education. Not specifically mathematical, but good general overview. MW

Education, P. *How to Start an Industrial Mathematics Program in the University.* Avner Friedman, John Lavery. SIAM, 1993, v + 37 pp, free (P). [ISBN 0-89871-327-7] For depart-

ments considering creating a graduate industrial mathematics program; based on experiences of the Institute for Mathematics and its Applications at the University of Minnesota. SM

History, L. *A History of Mathematical Notations: Two Volumes Bound as One.* Florian Cajori. Dover, 1993, \$19.95 (P). [ISBN 0-486-67766-4] V. I: *Notations in Elementary Mathematics*, xvi + 451 pp; V. II: *Notations Mainly in Higher Mathematics*, xii + 367 pp. For each symbol, details first appearance, origins, spread, and competition. Entertaining, interesting, good browsing. LC

Logic, P. *Complexity of Proofs and Their Transformations in Axiomatic Theories.* V.P. Orevkov. Transl. of Math. Mono., V. 128. AMS, 1993, vi + 153 pp, \$86. [ISBN 0-8218-4576-4]

Logic, P. *The Reconstruction of Trees from Their Automorphism Groups.* Matatyahu Rubin. Contemp. Math., V. 151. AMS, 1993, viii + 274 pp, \$56 (P). [ISBN 0-8218-5187-X]

Combinatorics, P. *Surveys in Combinatorics, 1993.* Ed: Keith Walker. London Math. Soc. Lect. Note Ser., V. 187. Cambridge Univ Pr, 1993, vii + 287 pp, \$39.95 (P). [ISBN 0-521-44857-3] 9 survey papers from the 14th British Combinatorial Conference (July 1993).

Discrete Mathematics, T(14-15: 1, 2). *Discrete Mathematics with Applications.* H.F. Mattson, Jr. Wiley, 1993, xxv + 637 pp, \$61.95. [ISBN 0-471-60672-3] Opening chapters on sets, logic, induction, equivalence relations and partitions, functions. Then a tree of possibilities: integers, congruences, binomial theorem, counting, probability, recurrence, matrices, trees, graphs. Conversational style, pleasing design. Many examples, problems of all kinds. Less algorithmic than other texts. LCL

Discrete Mathematics, T(16-17: 1), P*, L. *generatingfunctionology, Second Edition.* Herbert S. Wilf. Academic Pr, 1994, ix + 228 pp, \$44.95. [ISBN 0-12-751956-4] Excellent introduction to theory and techniques of generating functions. *Second Edition* has more problems, new applications, appendix on using Mathematica and Maple. (*First Edition*, TR and Extended Review, November 1990.) TH

Number Theory, T(16). *Primes and Programming: An Introduction to Number Theory with Computing.* Peter Gibling. Cambridge Univ Pr, 1993, x + 239 pp, \$44.95; \$19.95 (P). [ISBN 0-521-40182-8; 0-521-40988-8] An interesting, sophisticated introduction to number theory. Several Pascal programs, and many valuable computing exercises. Covers primes, congruences, pseudoprimes, cryptography, primitive

roots, divisors, continued fractions, quadratic residues. Have a look. SG

Number Theory, P. *Arithmetic of Quadratic Forms.* Yoshiyuki Kitaoka. Tracts in Math., V. 106. Cambridge Univ Pr, 1993, x + 268 pp, \$54.95. [ISBN 0-521-40475-4] A basic overview. After a general introduction to quadratic forms, systematically treats forms over the reals, rational local fields, p -adic integers, and the integers. A few exercises. SG

Linear Algebra, T(13-15: 1), L. *Linear Algebra: An Introduction to the Theory and Use of Vectors and Matrices.* Alan Tucker. Macmillan, 1993, xiii + 440 pp. [ISBN 0-02-421581-3] A new, refreshing look at elementary linear algebra. Reflects recent ideas on curricular reform and applications. Standard topics are enhanced by attractive examples, presentation. Scalar product, orthogonality are used systematically for matrix operations; eigenvalues, eigenvectors are introduced early. Well-selected exercises. JS

Algebra, T(13-14: 1), L. *Classical Algebra, Third Edition.* William J. Gilbert, Scott A. Vanstone. Waterloo Math Found (U. of Waterloo, Waterloo, Ontario, Canada N2L 3G1), 1993, vii + 248 pp, (P). [ISBN 0-921418-92-2] A curious, interesting, unusual book. Preparation for college-level modern algebra: elementary number theory, congruences, binomial theorem, permutation groups, cryptography, complex numbers, fundamental theorem of algebra (with intuitive topological proof), polynomial equations over finite and infinite fields. JS

Algebra, S(18), P. *Basic Structures of Modern Algebra.* Yuri Bahturin. Math. & Its Applic., V. 265. Kluwer Academic, 1993, ix + 419 pp, \$182. [ISBN 0-7923-2459-5] Best suited for expert reference; too terse, fast-paced for use as a beginning text. Topics include Galois theory, normed fields, simple groups, topological groups, Noetherian rings, central simple algebras, Lie algebras, homological algebras, algebraic groups, varieties of algebras. No examples or exercises. JS

Algebra, T(18). *Squares.* A.R. Rajwade. London Math. Soc. Lect. Note Ser., V. 171. Cambridge Univ Pr, 1993, xii + 286 pp, \$39.95 (P). [ISBN 0-521-42668-5] A well-written, unpretentious introduction to squares and sums of squares in fields. Among topics considered: the state of a field, Hilbert's 17th problem, Pfister forms, formally real fields. Some exercises. SG

Algebra, T(17: 2), P, L.** *Algebra: A Graduate Course.* I. Martin Isaacs. Brooks/Cole, 1994, xii + 516 pp, \$67.50. [ISBN 0-534-19002-2] A delicious, beautiful, infectiously

enthusiastic treatment of noncommutative and commutative algebra. Classical approach (no tensor products, categories), but includes modern topics. Nice examples, problems. RM

Algebra, T(16–17). *Finite Fields: Structure and Arithmetics.* Dieter Jungnickel. Bibliographisches Institut & FA Brockhaus, 1993, 339 pp, DM 78. [ISBN 3-411-16111-6] Explicit constructions, normal and dual bases, shift register sequences, characters and Gauss sums. Interesting commentaries; no exercises. SG

Algebra, P. *Computational Algebraic Geometry and Commutative Algebra.* Eds: David Eisenbud, Lorenzo Robbiano. Symp. Mathematica, V. XXXIV. Cambridge Univ Pr, 1993, x + 298 pp, \$49.95. [ISBN 0-521-44218-4] Papers from 1991 conference in Cortona on computational aspects of Gröbner bases. Includes tutorials, research surveys, open problems. RM

Algebra, T(18: 1, 2), P. *Quadratic Algebras, Clifford Algebras, and Arithmetic Witt Groups.* Alexander J. Hahn. Universitext. Springer-Verlag, 1994, xi + 286 pp, \$34 (P). [ISBN 0-387-94110-X] Main themes: algebras and forms over a commutative ring, involutions on algebras, gradings and tensor products, separable algebras. New results on representations of Clifford algebras, structure of the Arf algebra and the quadratic Witt group, connections between the group of quadratic algebras and the discriminant group. TH

Algebra, T(18: 1), P*.** *Hopf Algebras and Their Actions on Rings.* Susan Montgomery. CBMS Reg. Conf. Ser. in Math., No. 82. AMS, 1993, xiv + 238 pp, \$32 (P). [ISBN 0-8218-0738-2] Recent results on algebraic structure of Hopf algebras, their actions and coactions. Unifies theories of group, Lie algebra, and graded algebra actions. Provides an accessible introduction to quantum groups. TH

Calculus, T(13). *Calculus with Analytic Geometry.* Joe Repka. Wm C Brown, 1994, xxii + 1321 pp, \$75.40. [ISBN 0-697-06918-4] Not a lean book. Similar in coverage and emphasis to standard texts of the last 15 years. Readers may find the pages busy rather than lively: 4-color displays showing calculator keystrokes; historical asides; lots of colored graphics; 3-level problem sets with starred problems; calculator problems; and sometimes a "Point to Ponder." AWR

Calculus, P. *Differentialrechnung für Funktionen mit mehreren Variablen.* Klaus Harbarth, Thomas Riedrich, Winfried Schirotzek. BG Teubner Leipzig, 1993, 198 pp, DM 22,80 (P). [ISBN 3-8154-2041-5]

Calculus, P. *Integralrechnung für Funktionen*

mit mehreren Variablen. Karl-Heinz Körber, Ernst-Adam Pforr. BG Teubner Leipzig, 1993, 199 pp, DM 22,80 (P). [ISBN 3-8154-2042-3]

Calculus, T(13: 3, 4). *Applied Calculus, Second Edition.* Stanley I. Grossman. Wm C Brown, 1993, xix + 856 pp, \$61.26. [ISBN 0-697-11350-7] Introduction to calculus stressing applications to business, economics, social sciences, biology. Many solved examples, realistic applications, historical notes. KB

Calculus, C. *Insight: Demonstration Software for Calculus and Analytic Geometry, Eighth Edition, by Thomas and Finney.* IntelliPro. Addison-Wesley, 1992, iii + 6 pp, (P) with PC disk. [ISBN 0-201-90897-2] Graphics mostly well done. Good animation, but resolution is limited; some ε - δ pictures are hard to decipher. Many examples prompt user for parameter values, allowing experiments. Best used in careful coordination with classroom work. SM

Calculus, P. *Differential- und Integralrechnung für Funktionen mit einer Variablen.* Ernst-Adam Pforr, Winfried Schirotzek. BG Teubner Leipzig, 1993, 302 pp, DM 28,80 (P). [ISBN 3-8154-2040-7]

Real Analysis, T(18). *Measure Theory.* J.L. Doob. Grad. Texts in Math., V. 143. Springer-Verlag, 1994, xii + 210 pp, \$49. [ISBN 0-387-94055-3] A concise exposition. Measure spaces and measurable functions, integration and Hilbert space, measure sequences and signed measures, functions of bounded variation and Martingale theory. Probability notions are kept consistently before reader. No exercises (except to supply some missing proofs). AWR

Differential Equations, P. *Asymptotic Analysis: Linear Ordinary Differential Equations.* Mikhail V. Fedoryuk. Transl: Andrew Rodick. Springer-Verlag, 1993, viii + 363 pp, \$129. [ISBN 0-387-54810-6] Translation of 1983 Russian original. A comprehensive reference for asymptotic methods. SK

Differential Equations, T(18), P. *Introduction to Functional Differential Equations.* Jack K. Hale, Sjoerd M. Verduyn Lunel. Appl. Math. Sci., V. 99. Springer-Verlag, 1993, x + 447 pp, \$49. [ISBN 0-387-94076-6] Lively, thoughtful introduction. Motivates theory with applications (historical and recent), and the promise of a rich mathematical structure. Emphasizes dynamics, equations that depend on "past history." Well-written; extensive bibliography. MLR

Dynamical Systems, T(15–18: 1), S, L. *Frac-tals Everywhere, Second Edition.* Michael F. Barnsley. Academic Pr, 1993, xiv + 531 pp, \$49.95. [ISBN 0-12-079061-0] New features

include extensive exercise solutions, some new problems. A heady mix of analysis, geometry, dynamics, and computing, forcefully expounded and vividly illustrated. (*First Edition*, TR, April 1989; Extended Review, March 1990.) PZ

Numerical Analysis, T(16–17: 1), L. *Iterative Solution of Large Sparse Systems of Equations*. Wolfgang Hackbusch. Appl. Math. Sci., V. 95. Springer-Verlag, 1994, xxi + 429 pp, \$59. [ISBN 0-387-94064-2] Slightly revised translation from German. Treats Gauss-Seidel, conjugate gradient, multigrid, and domain decomposition methods. Examples from numerical PDE's. SM

Numerical Analysis, P. *Moving-grid Methods for Time-dependent Partial Differential Equations*. P.A. Zegeling. CWI Tract, V. 94. Centrum voor Wiskunde en Informatica, 1993, 168 pp, Dfl. 50 (P). [ISBN 90-6196-424-5]

Numerical Analysis, P, L. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. Richard Barrett, et al. SIAM, 1994, xiii + 112 pp, \$18 (P). [ISBN 0-89871-328-5] Provides templates (pseudocode descriptions of algorithms) to help with choice and implementation of methods effective on high-performance computers. AO

Numerical Analysis, P. *Numerical Methods for the Three-dimensional Shallow Water Equations on Supercomputers*. E.D. de Goede. CWI Tract, V. 88. Centrum voor Wiskunde en Informatica, 1993, 124 pp, Dfl. 40 (P). [ISBN 90-6196-417-2]

Numerical Analysis, P. *Numerical Linear Algebra*. Eds: Lothar Reichel, Arden Ruttan, Richard S. Varga. Walter de Gruyter, 1993, ix + 199 pp, DM 168. [ISBN 3-11-013784-4] Proceedings of a 1992 conference at Kent State University.

Operator Theory, T?(16–17: 1, 2), S, P. *Theory of Linear Operators in Hilbert Space: Two Volumes Bound As One*. N.I. Akhiezer, I.M. Glazman. Transl: Merlynd Nestell. Dover, 1993, xi + 218 pp, \$9.95 (P). [ISBN 0-486-67748-6] Unaltered republication of 1961 and 1963 editions. Bounded and unbounded operators; spectral analysis of completely continuous, unitary, and self-adjoint operators; extensions of symmetric operators. Appendices on differential operators and Naimark-Krein theory of generalized extensions of symmetric operators. No exercises. HD

Functional Analysis, P. *Evolutionary Integral Equations and Applications*. Jan Prüss. Mono. in Math., V. 87. Birkhäuser, 1993, xxvi + 366 pp, \$139. [ISBN 0-8176-2876-2]

Functional Analysis, P. *A Topological Introduction to Nonlinear Analysis*. Robert F. Brown. Birkhäuser, 1993, viii + 146 pp, \$24.50 (P). [ISBN 0-8176-3706-0] Topological approach to Krasnoselski-Rabinowitz bifurcation theorem and its application to Euler buckling of columns. SK

Functional Analysis, P. *Problèmes d'Analyse Fonctionnelle et d'Analyse Harmonique*. M. Samuelides, L. Touzillier. Cépaduès-Editions, 1993, vi + 391 pp. [ISBN 2-85428-257-4]

Analysis, P. *Recent Advances in Wavelet Analysis*. Eds: Larry L. Schumaker, Glenn Webb. Wavelet Analysis & Its Applic., V. 3. Academic Pr, 1994, xi + 364 pp. [ISBN 0-12-632370-4] Ten papers on theoretical and practical applications of wavelet analysis.

Analysis, T(15–17: 2), S**, L**.** *Advanced Calculus: A Differential Forms Approach*. Harold M. Edwards. Birkhäuser, 1994, xv + 508 pp, \$49.50. [ISBN 0-8176-3707-9] An inviting, unusual, high-level introduction to vector calculus, based solidly on differential forms. Superb exposition: informal but sophisticated, down-to-earth but general, geometrically and physically intuitive but mathematically rigorous, entertaining but serious. Remarkably diverse applications, physical and mathematical. Still fresh and novel at age 25. (1969 original edition, TR, October 1969; 1980 edition, TR, March 1981; Extended Review, March and December 1982.) PZ

Analysis, S(15–18), P, L. *Exercises for Fourier Analysis*. T.W. Körner. Cambridge Univ Pr, 1993, x + 385 pp, \$54.95; \$22.95 (P). [ISBN 0-521-43276-6; 0-521-43849-7] Chapter by chapter exercises for the author's splendid *Fourier Analysis* (TR, May 1990). KS

Analysis, T(17), S, P, L. *Wavelets: An Elementary Treatment of Theory and Applications*. Ed: Tom H. Koornwinder. Ser. in Approx. & Decompositions, V. 1. World Scientific, 1993, xii + 225 pp, \$48. [ISBN 981-02-1388-3] 12 papers from a 4-day course at CWI, Amsterdam. First 6 form nice introduction to (affine) wavelet theory; others treat applications, special aspects. KS

Algebraic Geometry, P. *Computational Algebraic Geometry*. Eds: Frédéric Eysssette, André Galligo. Progress in Math., V. 109. Birkhäuser, 1993, ix + 328 pp, \$64.50. [ISBN 0-8176-3678-1] 21 papers from the MEGA-92 symposium held in Nice, France.

Algebraic Geometry, P. *Cohomological Methods in Transformation Groups*. C. All-day, V. Puppe. Stud. in Adv. Math., V. 32.

Cambridge Univ Pr, 1993, xi + 470 pp, \$69.95. [ISBN 0-521-35022-0]

Algebraic Geometry, P. *Queen's Lectures on Arithmetical Composition of Quadratic Forms.* Anatolii N. Andrianov. Papers in Pure & Appl. Math., V. 92. Queen Univ, 1992, ii + 62 pp, (P). [ISBN 0-88911-634-2]

Geometry, T(15-16: 1, 2), L. *Modern Geometries, Fourth Edition.* James R. Smart. Brooks/Cole, 1993, xiii + 410 pp, \$51.50. [ISBN 0-534-21198-4] New edition has more examples and figures, short section on fractals, stronger bibliography, and more historical notes. (*First Edition*, TR, November 1973; Extended Review, August-September 1974.) DP

Algebraic Topology, P. *Morse Homology.* Matthias Schwarz. Progress in Math., V. 111. Birkhäuser, 1993, ix + 235 pp, \$49.50. [ISBN 0-8176-2904-1]

Algebraic Topology, P. *Nielsen Theory and Dynamical Systems.* Ed: Christopher K. McCord. Contemp. Math., V. 152. AMS, 1993, xii + 350 pp, \$52 (P). [ISBN 0-8218-5181-0] Proceedings of the 1992 AMS-IMS-SIAM Joint Summer Research Conference held at Mt. Holyoke College.

Topology, T(15-16: 1). *Topology of Surfaces.* L. Christine Kinsey. Undergrad. Texts in Math. Springer-Verlag, 1993, viii + 262 pp, \$39. [ISBN 0-387-94102-9] Choice of topics is "fairly random:" about equal parts point-set, combinatorial, and algebraic topology. SK

Topology, P*, L. *Knot Theory.* Charles Livingston. Carus Math. Mono., V. 24. MAA, 1993, xviii + 240 pp, \$31.50. [ISBN 0-88385-027-3] A lively exposition; requires only basic linear algebra. Covers (1) history and foundations of knot theory; (2) advanced topics: symmetry, Alexander polynomials, numerical invariants; (3) recent advances: the Conway, Jones, and Kauffman polynomials. TH

Operations Research, S(15-17), C, L. *Exploring Interior-Point Linear Programming: Algorithms and Software.* Ami Arbel. Found. of Comp. Ser. MIT Pr, 1993, xxiv + 211 pp, \$35 (P), with disk. [ISBN 0-262-51073-1] Treats 3 interior-point linear programming methods: primal affine scaling, dual affine scaling, and primal-dual methods. With excellent demonstration software (DOS) for problems with up to 100 constraints. No exercises. SM

Optimization, P, L. *Optimization Software Guide.* Jorge J. Moré, Stephen J. Wright. Frontiers in Appl. Math., V. 14. SIAM, 1993, xii + 154 pp, \$24.50 (P). [ISBN 0-89871-322-6] Notes for 1992 SIAM annual meeting short course. First part overviews al-

gorithms for different classes of optimization problems. Second part contains brief product descriptions from software vendors, individual researchers. AO

Optimization, T(18: 2), L. *Optima and Equilibria: An Introduction to Nonlinear Analysis.* Jean-Pierre Aubin. Grad. Texts in Math., V. 140. Springer-Verlag, 1993, xvi + 417 pp, \$59. [ISBN 0-387-52121-6] Convex optimization, 2- and n -person noncooperative games, cooperative games, nonlinear equations, variational inequalities, economic equilibria. Many exercises. Fully rigorous. SM

Mathematical Modeling, S(15-16). *Heureka heute: Kostproben praxiswirksamer Mathematik.* Günter Dewess, et al. BG Teubner Leipzig, 1993, 150 pp, DM 16,80 (P). [ISBN 3-8154-2071-7] Intended for a lay audience, but assumes a fair amount of mathematics. Each chapter treats a problem from real life (e.g., producing a chemical, routing coal trains, boring non-round holes). JD-B

Mathematical Modeling, T(13: 2), L. *Game Theory and Strategy.* Philip D. Straffin. New Math. Lib., V. 36. MAA, 1993, x + 244 pp, \$27.50 (P). [ISBN 0-88385-637-9] In-depth treatment of game theory. Assumes only high school algebra, but builds mathematical sophistication and familiarity with mathematical modeling process. Chapters on theory alternate with chapters on applications to anthropology, social psychology, economics, politics, business, biology, philosophy. Many exercises. KB

Control Theory, T?(17: 1), P. *Direct Adaptive Control Algorithms: Theory and Applications.* Howard Kaufman, Izhak Bar-Kana, Kenneth Sobel. Communications & Control Eng. Ser. Springer-Verlag, 1994, xxiii + 370 pp, \$69. [ISBN 0-387-94155-X] Theory and practice of adaptive control algorithms for multiple input/output control systems with parameter uncertainty. Applications include flexible structure control, drug infusion, robotics. RM

Control Theory, P. *Representation and Control of Infinite Dimensional Systems, Volume II.* Alain Bensoussan, et al. Systems & Control: Found. & Applic. Birkhäuser, 1993, xviii + 343 pp, \$79.50. [ISBN 0-8176-3642-0] Theory of quadratic cost optimal control for a large class of infinite-dimensional systems. JO

Control Theory, P. *Lecture Notes in Control and Information Sciences-189: Non-Identifier-Based High-Gain Adaptive Control.* Achim Ilchmann. Springer-Verlag, 1993, x + 204 pp, \$54 (P). [ISBN 0-387-19845-8]

Probability, P. *Ten Lectures on the Probabilistic Method, Second Edition.* Joel Spencer.

CBMS-NSF Reg. Conf. Ser. in Appl. Math., V. 64. SIAM, 1994, vi + 88 pp, \$18.50 (P). [ISBN 0-89871-325-0] Based on lectures given at 1986 CBMS-NSF conference held at Fort Lewis College. This edition contains new results, additional material. (*First Edition*, TR, December 1988.) LC

Mathematical Statistics, P. *Efficient and Adaptive Estimation for Semiparametric Models.* Peter J. Bickel, *et al.* Ser. in Math. Sci. Johns Hopkins Univ Pr, 1993, xix + 560 pp, \$95. [ISBN 0-8018-4541-6] Asymptotic inference for finite-dimensional parametric models, with extensions to semiparametric models. Information bounds for infinite-dimensional parameters, construction of estimates. Limited results on general methods for constructing asymptotically efficient estimates. RWJ

Statistical Methods, P. *Statistical Uncertainties in Posterior Probabilities.* A.W. Ambergen. CWI Tract, V. 93. Centrum voor Wiskunde en Informatica, 1993, 129 pp, Dfl. 40 (P). [ISBN 90-6196-422-9]

Programming, P. *Mastering C Pointers: Tools for Programming Power, Second Edition.* Robert J. Traister. Academic Pr, 1993, xii + 163 pp, (P), with disk. [ISBN 0-12-697409-8]

Languages, S(17), P, L. *The High Performance Fortran Handbook.* Charles H. Koebel, *et al.* Scientific & Eng. Comput. MIT Pr, 1994, xiv + 329 pp, \$24.95 (P); \$45. [ISBN 0-262-61094-9; 0-262-11185-3] High Performance Fortran, an extension of Fortran 90, is a portable programming language that allows easy access to vector and massively parallel multiprocessors. Book presents HPF in tutorial form. SM

Computer Systems, P. *Learning Perl.* Randal L. Schwartz. O'Reilly & Assoc, 1993, xxv + 246 pp, \$24.95 (P). [ISBN 1-56592-042-2]

Computer Systems, S(15), C, L.** *Solving Problems in Scientific Computing Using Maple and MATLAB.* Walter Gander, Jiří Hřebíček. Springer-Verlag, 1993, xiii + 268 pp, \$39 (P). [ISBN 0-387-57329-1] 19 nifty examples of high-level classroom applications of Maple and MATLAB: trajectories of tennis balls, orbits in the three-body problem, smoothing filters, conformal mappings, heat flow, and the compression of a metal disk. An excellent reference on undergraduate mathematical computing. MPR

Computer Science, P. *Statistical Analysis of Software Reliability Models.* M.C.J. van Pul. CWI Tract, V. 95. Centrum voor Wiskunde en Informatica, 1993, 186 pp, Dfl. 50 (P). [ISBN 90-6196-425-3]

Applications (Biological Science), P. *Moment Problems in Hilbert Space with Applications*

to Magnetic Resonance Imaging. M. Zwaan. CWI Tracts, V. 89. Centrum voor Wiskunde en Informatica, 1993, xv + 136 pp, Dfl. 40 (P). [ISBN 90-6196-418-0]

Applications (Physics), T(15-16: 1, 2), S, L. *Spacetime Without Reference Frames.* Tamás Matolcsi. Akadémiai Kiadó, 1993, 411 pp, \$50. [ISBN 963-05-6433-5] This ambitious book proceeds quickly from easy introductory generalities to sophisticated mathematical physics. Underlying mathematical concepts (tensor analysis, manifolds, and Lie groups) occupy last 100 pages. MU

Applications (Physics), S, P. *Quantum Inverse Scattering Method and Correlation Functions.* V.E. Korepin, N.M. Bogoliubov, A.G. Izergin. Mono. on Math. Physics. Cambridge Univ Pr, 1993, xix + 555 pp, \$100. [ISBN 0-521-37320-4] An important, carefully-crafted text, in 4 parts: examination of the Bethe ansatz and calculation of physical quantities; theory of the quantum inverse scattering; third and fourth sections apply preceding work to calculation of correlation functions. MU

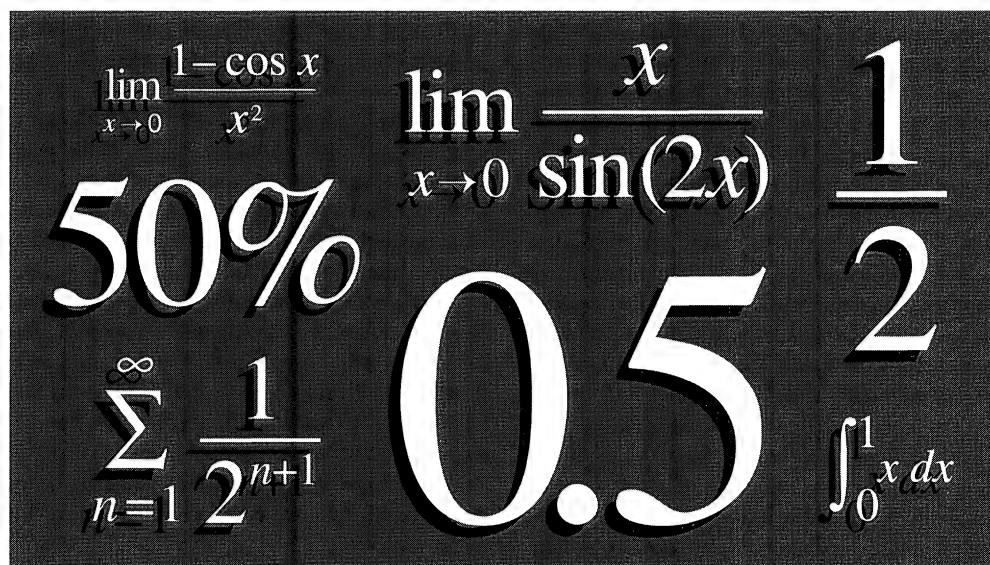
Applications (Physics), S(15-17), P, L. *The Philosophy Behind Physics.* Thomas Brody. Eds: Luis de la Peña, Peter E. Hodgson. Springer-Verlag, 1993, xii + 355 pp, \$59. [ISBN 0-387-55914-0] This masterpiece should be much appreciated by physicists, but of less value to philosophers. Author's technical expertise is obvious throughout; thought-provoking text is virtually devoid of esoteric philosophical vocabulary, but rich in precise examples, lucid discussion, and analysis. MU

Applications (Simulation), P. *Algebraic Specification of Communication Protocols.* Eds: S. Mauw, G.J. Veltink. Tracts in Theoret. Comp. Sci., V. 36. Cambridge Univ Pr, 1993, xi + 197 pp, \$39.95. [ISBN 0-521-41883-6] Surveys Process Specification Formalism (PSF), a formal method for specifying and testing unambiguous mathematical models. Chapters range from tutorial on PSF through research topics. RM

Reviewers

KB: Karla Ballman, Macalester; LC: Laura Chihara, St. Olaf; HD: Hung Dinh, Macalester; JD-B: John Dyer-Bennet, Carleton; SG: Steven Galovich, Carleton; TH: Tom Halverson, Macalester; RWJ: Roger W. Johnson, Carleton; SK: Steve Kennedy, St. Olaf; LCL: Loren C. Larson, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; JO: Jeff Ondich, Carleton; AO: Arnold Ostebee, St. Olaf; DP: David Peifer, St. Olaf; MLR: Margaret L. Reese, St. Olaf; MPR: Matthew P. Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; KS: Karen Saxe, Macalester; JS: John Schue, Macalester; MU: Milton Ulmer, Carleton; MW: Martha Wallace, St. Olaf; PZ: Paul Zorn, St. Olaf.

**No matter how you
express it, it still means
DERIVE® is half price.**



DERIVE →

The *DERIVE A Mathematical Assistant* program lets you express yourself symbolically, numerically and graphically, from algebra through calculus, with vectors and matrices too—all displayed with accepted math notation, or 2D and 3D plotting. *DERIVE* is also easy to use and easy to read, thanks to a friendly, menu-driven interface and split or

overlay windows that can display both algebra and plotting simultaneously. Better still, *DERIVE* has been praised for the accuracy and exactness of its solutions. But, best of all the suggested retail price is now only \$125. Which means *DERIVE* is now half price, no matter how you express it.

System requirements

DERIVE: MS-DOS 2.1 or later, 512K RAM, and one 3½" disk drive. Suggested retail price now **\$125 (Half off!)**.

DERIVE ROM card: Hewlett Packard 95LX & 100LX Palmtop, or other PC compatible ROM card computer. Suggested retail price now **\$125!**

DERIVE XM (eXtended Memory): 386 or 486 PC compatible with at least 2MB of *extended* memory. Suggested list price now \$250!

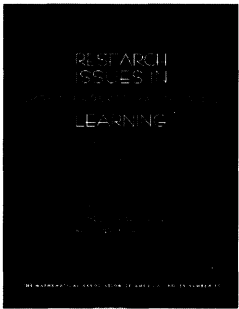
DERIVE is a registered trademark of Soft Warehouse, Inc.

 **Soft Warehouse**
HONOLULU • HAWAII

Soft Warehouse, Inc. • 3660 Waiālae Ave.
Ste. 304 • Honolulu, HI, USA 96816-3236
Ph: (808) 734-5801 • Fax: (808) 735-1105

Research Issues in Undergraduate Mathematics Learning Preliminary Analyses and Reports

James J. Kaput and Ed Dubinsky, Editors



Research in undergraduate mathematics education is important for all college and university mathematicians. If our students are to be more successful in understanding mathematics, then college faculty need to understand how mathematics is learned. This knowledge can guide us in curriculum reform and in improving our own teaching. It can help us make mathematics accessible to all students and it can increase the number of graduate students in mathematics.

This volume of research in undergraduate mathematics education informs us about the nature of student learning in some of the most important topics in the undergraduate curriculum: sets, functions, calculus, statistics, abstract algebra and problem solving. Paying careful attention to the trouble students have in learning mathematics will help us to work with students so they can deal with those difficulties.

A survey of the literature begins the volume. Becker and Pence have brought together an unusually complete list of references on research in collegiate mathematics. Their comments will guide those attempting to begin or to continue a program of research in student learning.

The sad fact that even good calculus students stumble over nonroutine problems is the theme of Selden, Selden, and Mason. Their conclusions point to significant shortcomings in the curriculum. This study of student difficulties is

continued by Ferrini-Mundy and Graham who investigate a single student's interactions with the fundamental concepts of the calculus. Baxter studies a group of students to learn how they acquire the concept of set, while Cuoco does the same for the concept of function.

Cooperative learning does help the student. That is the conclusion of Bonsangue, who investigates how two carefully matched classes of students in a statistics course perform on exams. How students learn to write proofs in group theory is the subject considered by Hart. Rosamond breaks new ground by comparing how emotions vary in their effect on the problem solving ability of novices and experts.

All college faculty should read this book to find how they can help their students learn mathematics.

150 pp., Paperbound, 1994
ISBN 0-88385-090-7

List: \$24.00
Catalog Number NTE-33

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Membership Code _____

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty.	Catalog Number	Price
_____	_____	_____
_____	_____	_____
		Total \$ _____
Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
Credit Card No. _____		
Signature _____		
Exp. Date _____		

Essential Mathematics from Cambridge

Probability Theory, an Analytic View

Daniel W. Stroock

Presents useful material for students who, even if they do not intend to devote a major portion of their careers to the study of probability theory, want to know what they are missing if they do not.

1994 527 pp. 43123-9 Hardback \$49.95

How To Prove It

A Structured Approach

Daniel J. Velleman

Prepares students to make the transition from solving problems to proving theorems by teaching the techniques needed to read and write proofs.

1994 304 pp. 44116-1 Hardback \$49.95
44663-5 Paperback \$19.95

Elementary Probability

David Stirzaker

Illustrates the wide range and power of the subject by including conditional probability, independence, random variables, generating functions, and an introduction to Markov chains.

1994 416 pp. 42028-8 Hardback \$64.95
42183-7 Paperback \$24.95

Representations and Characters of Groups

Gordon James

and Martin Liebeck

Provides a modern introduction to the representation theory of finite groups.

1993 429 pp. 44024-6 Hardback \$69.95
44590-6 Paperback \$29.95

Basic Abstract Algebra

Second Edition

**P. B. Bhattacharya, S. K. Jain,
and S. R. Nagpaul**

"...a thorough and surprisingly clean-cut survey of the group/ring/field troika which manages to convey the idea of algebra as a unified enterprise."

—**Ian Stewart, New Scientist**

1994 416 pp. 46081-6 Hardback \$69.95
46629-6 Paperback \$29.95

Categories for Types

Roy L. Crole

Explains the basic principles of categorical type theory and the techniques used to derive categorical semantics for specific type theories.

1994 352 pp. 45092-6 Hardback \$59.95
45701-7 Paperback \$27.95

Computational Geometry in C

Joseph O'Rourke

Covers the basic techniques used in computational geometry—polygon triangulations, convex hulls, Voronoi diagrams, and arrangements.

1994 327 pp. 44034-3 Hardback \$59.95
44592-2 Paperback \$24.95

Now in paperback...

Numbers and Functions

Steps to Analysis

R. P. Burn

"...invaluable as a teaching resource, or for independent study with gifted students."

—**The American
Mathematical Monthly**

346 pp. 45773-4 Paperback \$24.95

The Banach-Tarski Paradox

Stan Wagon

"...packed with fascinating and beautiful results." —**R. J. Gardner, Bulletin of
the London Mathematical Society**

"...this beautiful book is written with care and is certainly worth reading."

—**Włodzimierz Bzyl,
Mathematical Reviews**

271 pp. 45704-1 Paperback \$24.95

Available in bookstores or from

CAMBRIDGE
UNIVERSITY PRESS

40 West 20th St., N.Y., NY 10011-4211

Call toll-free 800-872-7423

MasterCard/VISA accepted.

Prices subject to change.

NEW IN THE SPECTRUM SERIES

Complex Numbers and Geometry

Liang-shin Hahn

The purpose of this book is to demonstrate that complex numbers and geometry can be blended together beautifully, resulting in easy proofs and natural generalizations of many theorems in plane geometry—such as the Napoleon theorem, the Ptolemy–Euler theorem, the Simson theorem, and the Morley theorem.

Beginning with a construction of complex numbers, readers are taken on a 140-page guided tour that includes something for everyone, even those with advanced degrees in mathematics. Yet, the entire book is accessible to a talented high-school student.

The book is self-contained—no background in complex numbers is assumed—and can be covered at a leisurely pace in a one-semester course. Many of the chapters can be read independently. Over 100 exercises are included. The book would be suitable as a text for a geometry course, or for a problem solving seminar, or as enrichment for the student who wants to know more.

200 pp., Paperbound, 1994

ISBN 0-88385-510-0

List: \$25.50 MAA Member: \$19.50

Catalog Number CNGE

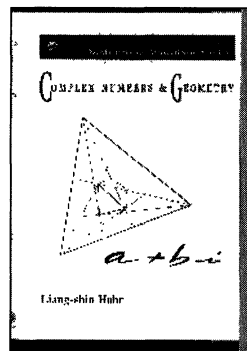


Table of Contents

1. Complex Numbers

Introduction to Imaginary Numbers; Definition of Complex Numbers; Quadratic Equations; Significance of the Complex Numbers; Order Relation in the Complex Field; The Triangle Inequality; The Complex Plane; Polar Representation of Complex Numbers; The n th Roots of 1; The Exponential Functions; Exercises

2. Applications to Geometry

Triangles; The Ptolemy–Euler Theorem; The Clifford Theorems; The Nine-Point Circle; The Simson Line; Generalizations of the Simson Theorem; The Cantor Theorems; The Feuerbach Theorem; The Morley Theorem; Exercises

3. Möbius Transformations

Stereographic Projection; Möbius Transformations; Cross Ratios; The Symmetry Principle; A Pair of Circles; Pencils of Circles; Fixed Points and the Classification of Möbius Transformations; Inversions; The Poincaré Model of a Non-Euclidean Geometry; Exercises

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty. Catalog Number Price

Total \$

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

Knot Theory

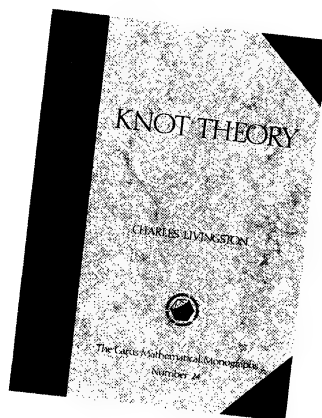
Charles Livingston

I learned more about knots after an hour with the book than I thought I could, and I am glad that it is here on my desk so that I may spend more time with it and, I hope, learn more.
—Paul Halmos

Knot Theory, a lively exposition of the mathematics of knotting, will appeal to a diverse audience from the undergraduate seeking experience outside the traditional range of studies to mathematicians wanting a leisurely introduction to the subject. Graduate students beginning a program of advanced study will find a worthwhile overview, and the reader will need no training beyond linear algebra to understand the mathematics presented.

Over the last century, knot theory has progressed from a study based largely on intuition and conjecture into one of the most active areas of mathematical investigation. **Knot Theory** illustrates the foundations of knotting as well as the remarkable breadth of techniques it employs—combinatorial, algebraic, and geometric.

The interplay between topology and algebra, known as algebraic topology, arises early in the book, when tools from linear algebra and from basic group theory are introduced to study the properties of knots, including the unknotting number, the braid index, and the bridge number. Livingston guides you through a general survey of the topic showing how to use the techniques of linear algebra to address some sophisticated problems, including one of mathematics' most beautiful topics, symmetry. The book closes with a discussion of high-dimensional knot theory and a presentation of some



of the recent advances in the subject—the Conway, Jones and Kauffman polynomials. A supplementary section presents the fundamental group, which is a centerpiece of algebraic topology.

An extensive collection of exercises is included. Some problems focus on details of the subject matter; others introduce new examples and topics illustrating both the wide range of knot theory and the present borders of our understanding of knotting. All are designed to offer the reader the experience and pleasure of working in this fascinating area.

264 pp., Hardbound, 1993

ISBN 0-88385-027-3

List: \$31.50 MAA Member: \$25.00

Catalog Number CAM-24

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036
1-(800) 331-1622 (202)-387-5200



Membership Code

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty. Catalog Number Price

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

Joby Milo Anthony, Editor

The conference was unique. Conference participants included pre-college mathematics teachers, community college and university teachers, and research mathematicians. Papers were delivered in sessions devoted to the classroom teacher, to the history of mathematics, and to pedagogical and research interests in geometry. Many lectures combined these subjects. This book presents some of those lectures. Anyone involved with teaching or producing mathematics can find something in this volume that will be interesting to them.

Some of these papers are specifically for the classroom teacher. They discuss a use of technology, or the organization of a class for some specific purpose. Other articles will provide teachers with examples of mathematical problems or historical episodes that can be used as part of a mathematics class. Still other papers deal with the



Also included in this volume is a penetrating interview with Eves.

220 pp., 1994, Paperbound

ISBN-0-88385-088-5

List: \$24.00

Catalog Number NTE-34

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Membership Code

Name _____

Address

City

State Zip Code

Qty.	Catalog Number	Price
------	----------------	-------

Total \$

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No.

Signature _____

Exp. Date _____

The Search for E.T. Bell

also known as John Taine

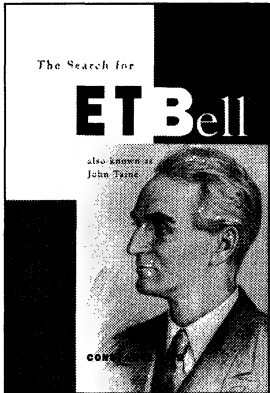
Constance Reid

No one today writes about mathematics and mathematicians with more grace, knowledge, skill, and clarity, and no one is going to produce a more delightful, informative, accurate account of Eric Temple Bell and his work, and that of his alter-ego, the prolific pioneer of science fiction, John Taine. This is a fine book. —Martin Gardner

Eric Temple Bell has been one of my heroes for 60 years...I congratulate Constance Reid on a remarkable achievement. I hope it is greeted with the success it deserves, and revives interest in an extraordinary and multi-talented man. —A. C. Clarke

Eric Temple Bell (1883–1960) was a distinguished mathematician and a best selling popularizer of mathematics. His *Men of Mathematics*, still in print after almost sixty years, inspired scores of young readers to become mathematicians. Under the name “John Taine,” he also published science fiction novels (among them *The Time Stream*, *Before the Dawn*, and *The Crystal Horde*) that served to broaden the subject matter of that genre during its early years.

In *The Search for E.T. Bell*, Constance Reid has given us a compelling account of this complicated, difficult man who never divulged to anyone, not even to his wife and son, the story of his early life and family background. Her book is thus more of a mystery than a traditional biography. It begins with the discovery of an unexpected inscription in an English churchyard and a series of cryptic notations in a boy’s schoolbook. Then comes an inadvertent revelation, by Bell himself, in a respected mathematical journal...You will have to read the book to learn the rest.



Originally agreeing to write only a profile of Bell, Mrs. Reid soon found herself involved in a full-length biography. The discoveries she made in the course of her five years of research will necessitate a fresh evaluation of his extensive mathematical work and his science fiction novels as well as the revision of almost every statement currently in print about his family background and early life. Mrs. Reid is already well known as the author of acclaimed biographies of David Hilbert, Richard Courant, and Jerzy Neyman.

Includes a collection of over 75 photographs.

384 pp., Hardbound, 1993

ISBN 0-88385-508-9

List: \$35.00 MAA Member: \$28.00

Catalog Number BELL

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
(202) 387-5200 1-(800) 331-1622

Membership Code -----	Qty. _____	Catalog Number _____	Price _____
Name _____	Total \$ _____		
Address _____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
City _____	Credit Card No. _____		
State _____ Zip Code _____	Signature _____		
	Exp. Date _____		

EXCURSIONS IN CALCULUS: an Interplay of the Continuous and the Discrete

Robert M. Young

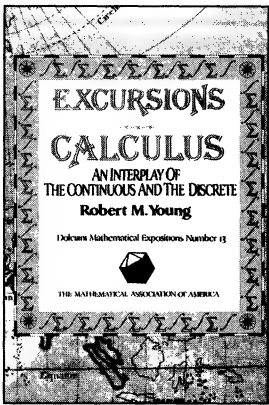
An excellent source of projects for well motivated students. This list of 463 references is a valuable aid for those who wish to dig deeper. —CHOICE

The presentation is clear and the topics very interesting...fully accessible to students for whom the book is intended. The book will be influential in awakening students' awareness for good classical mathematics. —Paulo Ribenboim

Printed with eight full-color plates.

The purpose of this book is to explore, within the context of elementary calculus, the rich and elegant interplay that exists between the two main currents of mathematics, the continuous and the discrete. Such fundamental notions in discrete mathematics as induction, recursion, combinatorics, number theory, discrete probability, and the algorithmic point of view as a unifying principle are continually explored as they interact with traditional calculus. The interaction enriches both.

The book is addressed primarily to well-trained calculus students and their teachers, but it can serve as a supplement in a traditional calculus course for anyone who wants to see more.



CONTENTS:

- Infinite Ascent, Infinite Descent: The Principle of Mathematical Induction
- Patterns, Polynomials, and Primes: Three Applications of the Binomial Theorem
- Fibonacci Numbers: Function and Form
- On the Average
- Approximation: from Pi to the Prime Number Theorem
- Infinite Sums: A Potpourri

The problems, taken for the most part from probability, analysis and number theory, are an integral part of the text. Many point the reader toward further excursions. There are over 400 problems presented in this book.

408 pp., 1992, Paperbound
ISBN 0-88385-317-5
List: \$42.00 MAA Member \$34.00
Catalog Number DOL-13

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

Membership Code	Qty.	Catalog Number	Price
Name _____	_____	_____	_____
Address _____	_____	_____	_____
City _____	_____	_____	Total \$ _____
State _____ Zip Code _____	_____	_____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
	_____	_____	Credit Card No. _____
	Signature _____	_____	Exp. Date _____

Visualization in Teaching and Learning Mathematics

Walter Zimmermann and
Steve Cunningham, Editors

Buy this book. If you can't buy it, have the library order it. If the library won't order it, ask to borrow a copy from a friend. But do read this book.

—*The Mathematics Teacher*

High school, community college, and university teachers who use or are interested in using graphics to teach calculus, deductive reasoning, functions, geometry, or statistics will find valuable ideas for teaching... A must for every college or university library with a mathematics department.—*CHOICE*

The twenty papers in this book give an overview of research, analysis, practical experience, and informed opinion about the role of visualization in teaching and learning mathematics, especially at the undergraduate level. Visualization in its broadest sense is as old as mathematics, but

progress in computer graphics has generated a renaissance of interest in visual representations and visual thinking in mathematics.

230 pp., Paperbound, 1991

ISBN 0-88385-071-0

List: \$24.00

Catalog Number NTE-19

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Two New Derive® Books Now Available

Numerical Analysis Via Derive

Steven Schonefeld

This book is written as a primary text for an introductory course in numerical analysis. For convenience, a disk is included with a number of the functions that are defined. The text is intended to permit the exploration of many different numerical algorithms with only a minimum of technical effort. In order to use this book, all you need is a computer and a copy of Derive (version 2.09, or higher). The computer disk that comes with this book contains all the Derive procedures used in the book. You do not need to be a computer programmer to work the exercises. Simply transfer the appropriate files into Derive and start using it. (529 pages Paperbound)
List \$44.95 MAA member (\$40.45)

Differential Equations With Derive

David C. Arney

(for Derive Version 2.5 and above)

This book is designed to show how to use Derive to help solve problems in Differential Equations and related subjects. It is a companion to any of the textbooks used in differential equations courses or related subject courses (i.e., engineering mathematics, applied mathematics, dynamical systems). Topics include: First Order Differential Equations, Numerical Methods and Difference Equations, Second and Higher Order Equations, Matrix Algebra and Systems of Equations, and Partial Differential Equations. (284 pages, Paperbound)

List Price \$22.95 MAA member (\$20.65)

Order From

MathWare 604 E. Mumford Dr. Urbana, IL 61801

(800)255-2468/ (217)384-3196 or Fax (217)384-7043



Game Theory and Strategy

Philip D. Straffin, Jr.



This valuable addition to the New Mathematical Library series pays careful attention to applications of game theory in a wide variety of disciplines. The applications are treated in considerable depth. The book assumes only high school algebra, yet gently builds to mathematical thinking of some sophistication. **Game Theory and Strategy** might serve as an introduction to both axiomatic mathematical thinking and the fundamental process of mathematical modelling. It gives insight into both the nature of pure mathematics, and the way in which mathematics can be applied to real problems.

Since its creation by John von Neumann and Oskar Morgenstern in 1944, game theory has contributed new insights to business, politics, economics, social psychology, philosophy, and evolutionary biology. In this book, the fundamental ideas of game theory share the stage with applications of the theory. How might strategic business decisions depend on information about a rival company, and how much would such information be worth? When is it advantageous to vote for a candidate who is not your favorite? What are the optimal strategies for teams in the football draft, and what paradoxes can result from following

those strategies? What is a fair way to share the costs of a development project? What can we learn about the problem of "free will" by imagining playing a game with an omnipotent Being? How might natural selection lead to altruistic behavior in animal species? Game theory gives insight into all of these questions.

The book includes many exercises, with answers, which allow the reader to try out calculations, and explore alternative formulations of game-theoretic ideas.

200 pp., 1993, Paperbound

ISBN 0-88385-637-9

List: \$27.50 MAA Member: \$22.00

Catalog Number NML-36

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
(202) 387-5200 Fax (800) 331-1622

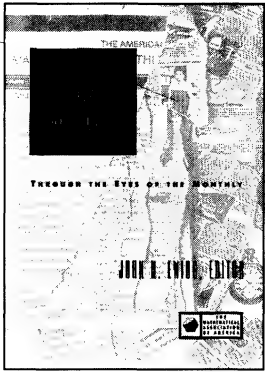
Membership Code -----	Qty.	Catalog Number	Price
Name _____			
Address _____			Total \$ _____
City _____			Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
State _____ Zip Code _____			Credit Card No. _____
			Signature _____
			Exp. Date _____

A Century of Mathematics Through the Eyes of the Monthly

John Ewing, Editor

This is the story of American mathematics during the past century. It contains articles and excerpts from a century of the **Monthly**, giving the reader and opportunity to skim all one hundred volumes without actually opening them. It samples mathematics year by year and decade by decade. Along the way, readers can glimpse the mathematical community at the turn of the century, and the divisions between the mathematical communities of teachers and researchers. They read about the struggle to prevent colleges from eliminating mathematics requirements in the 1920's, the controversy about Einstein and relativity, the debates about formalism in logic, the immigration of mathematicians from Europe, and the frantic effort to organize as the war began. At the end of the war, they hear about new divisions between pure and applied mathematics, heroic efforts to deal with large numbers of new students in the universities, and the rise of federal funding for mathematics. In more recent times, they see the advent of computers and computer science, the problems faced by women and minorities, and some of the triumphs of modern research.

This is a book about mathematics—about teaching and research, applied and pure, elite universities and community colleges. Browsing through its pages, readers see what has changed (the kinds of mathematics in fashion, for example) and what has stayed the same (our concern about teaching and our complaints about the deplorable state of our students).



This is a book about history—a sampling of history, meant to be savored rather than studied. For one hundred years, the **Monthly** has contained articles by some of the greatest mathematicians in the world, as well as articles by students and faculty from small midwestern colleges where those great names were barely known. This book gives a glimpse of both worlds. It tells a story rather than the details of history.

This is the story of a century of mathematics in America.

335 pp., Hardbound, 1994

ISBN 0-88385-457-0

List: \$39.50 MAA Member: \$32.00

Catalog Number CENTMA

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

	Qty.	Catalog Number	Price
Name _____			
Address _____			
City _____			
State ____ Zip Code _____			
			Total \$ _____
			Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
			Credit Card No. _____
			Signature _____
			Exp. Date _____

NEW IN THE SPECTRUM SERIES

Cryptology

Albrecht Beutelspacher

FR BRX XQFHUVWCQG WKLV? If you can't decipher this coded message, you must read this book!

How can messages be transmitted secretly? How can one guarantee that the message arrives safely in the right hands exactly as it was transmitted? Cryptology—the art and science of “secret writing”—provides ideal methods to solve these problems of data security.

Technology advances have stimulated interest in the study of cryptology. Of course, computers can break cryptosystems much more efficiently than humans can. Computers allow complex and sophisticated mathematical techniques which achieve a degree of security undreamt of by previous generations. Today the applications of cryptology range from the encryption of television programs sent via satellite, to user authentication of computers, to new forms of electronic payment systems using smart cards.

The first half of the book studies and analyzes classical cryptosystems. Here we find Caesar's cipher, the Spartan scytale, the Vigenère cipher, and more. The theory of cipher systems is presented, including a description of the best possible cipher, the one-time pad. An introduction to linear shift registers, which serve as

building blocks for most presently used ciphers, is also given.

The second half of the book looks at the exciting new directions of public-key cryptology, which since its invention in 1976, has revolutionized data security. The author also looks at the famous RSA-algorithm, algorithms based on “discrete logarithms,” the so-called zero-knowledge algorithms, and the smart cards that bring cryptographic services to the man-on-the-street.

Although the mathematics covered is nontrivial, the book is fun to read, and the author presents the material clearly and simply. Many exercises and references accompany each chapter. The book will appeal to a wide audience including teachers, students, and the interested layman.

Cryptology was originally published in German by Vieweg. This edition has been extensively revised.

176 pp., Paperbound, 1994

ISBN 0-88385-504-6

List: \$26.00 MAA Member: \$20.00

Catalog Number CRYPT

Name _____

Address _____

City _____

State _____ Zip Code _____

Qty.	Catalog Number	Price
------	----------------	-------

Total \$ _____

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

FROM MARTIN GARDNER

Mathematical Magic Show

...We visit most of the prime sites of recreational mathematics: game theory, factorials, puzzles, playing cards, finger arithmetic, Möbius bands, polyominoes, perfect numbers, the knight's tour, trees, and dice. Gardner always has new facts and ideas to add interest to even the most well-trodden areas. —Times Literary Supplement

312 pp., Paperbound, 1990
ISBN 0-88385-449-X
List: \$19.50 MAA Member: \$16.50
Catalog Number MAGIC

Mathematical Carnival

His startling gift for bringing the sublime to the people is unabated. Once again, hard mathematical ideas are conveyed with fluency, charm, and utter clarity. As a philosopher, I warmly salute his judgement of when to leave a metaphysical question enticingly open. His craftsmanship remains exquisite. — New Scientist

320 pp., Paperbound, 1988
ISBN 0-88385-448-1
List: \$18.00 MAA Member: \$15.00
Catalog Number MCR

Riddles of the Sphinx and Other Mathematical Puzzle Tales

This book charms, informs, inspires, puzzles, and delights, and the reader can dip in almost anywhere and get hooked by the natural lucidity of style and the friendly tone which are so characteristic of Martin Gardner.
—Mathematical Spectrum

184 pp., Paperbound, 1987
ISBN 0-88385-632-8
List: \$16.00 MAA Member: \$13.00
Catalog Number NML-32

Mathematical Circus

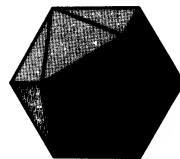
A circus suggests fun and enjoyment and there is plenty of both to be found here. The book should certainly be in the school library. It will also be valuable resource for the teacher.
—The Mathematical Gazette

300 pp., Paperbound, 1992
ISBN 0-88385-506-2
List: \$19.50 MAA Member: \$16.50
Catalog Number CIRCUS

ORDER FROM:
The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Membership Code	Qty.	Catalog Number	Price

Name _____			
Address _____			
City _____			Total \$ _____
State _____ Zip Code _____			Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
			Credit Card No. _____
			Signature _____ Exp. Date _____



Contents

ARTICLES

- A Tale of Two CD's / DAN KENNEDY 603
- Three Problems in Search of a Measure / JONATHAN L. KING 609
- The n -Queens Problem / IGOR RIVIN, ILAN VARDI, and
PAUL ZIMMERMANN 629
- What's the Difference Between Cantor Sets? / ROGER L. KRAFT 640
- Morphisms, Squarefree Strings, and the Tower of Hanoi Puzzle /
JEAN-PAUL ALLOUCHE, DAN ASTOORIAN, JIM RANDALL, and
JEFFREY SHALLIT 651

FEATURES

COMMENTS 602

NOTES

- Sierpinski's Theorem Is Deducible from Euler and Dirichlet /
A. A. AGEEV 659
- On Nonnegativity of Symmetric Polynomials / F. MATÚŠ 661
- New Tricks for Old Trees: Maps and the Pigeonhole Principle /
N. GRAHAM, R. C. ENTRINGER, AND L. A. SZÉKELY 664

THE COMPUTER SCIENCE SAMPLER

- Do You Know the Way to Vertex A ? / JEFF ONDICH 668

THE EVOLUTION OF...

- On the Calculus of Variations and Its Major Influences
on the Mathematics of the First Half of Our Century. Part I. /
ERWIN KREYSZIG 674

THE AUTHORS 679

PROBLEMS AND SOLUTIONS 681

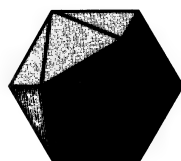
REVIEWS

- How to Teach Mathematics*, By Steven G. Krantz /
MEYER JERISON 692

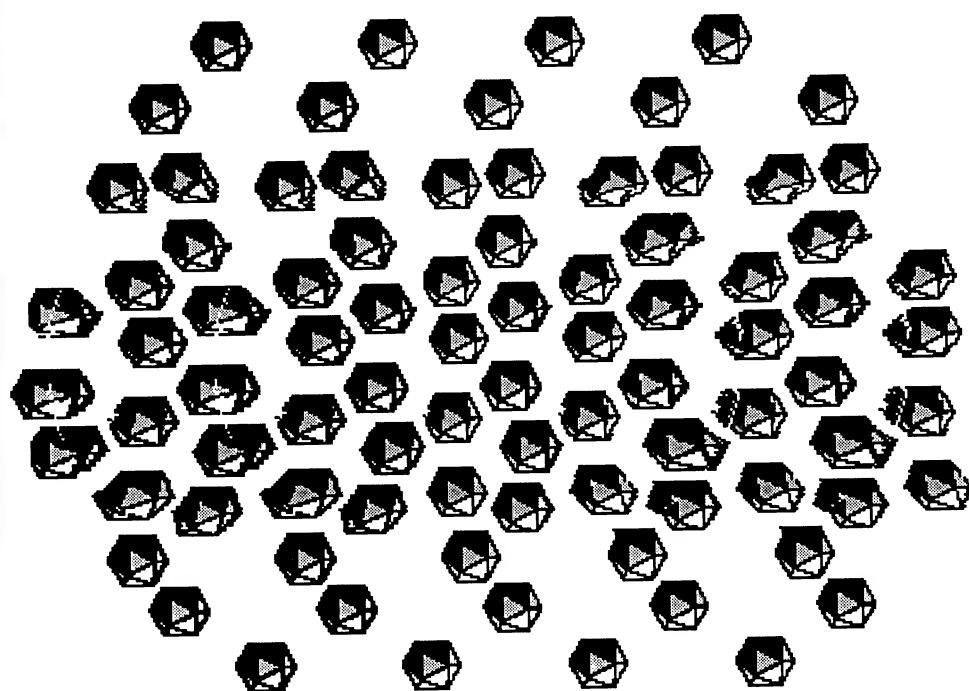
TELEGRAPHIC REVIEWS 695



The American Mathematical Monthly



Volume 101, Number 8 / OCTOBER 1994



NOTICE TO AUTHORS

The *Monthly* publishes articles, notes, and other features about mathematics and the profession. The readership of the *Monthly* is intended to include everybody who is mathematically inclined, including of course professional mathematicians and students of mathematics at all collegiate levels. While no single article or feature is likely to appeal to everyone, material should interest and be accessible to a large number of readers. This is the most important criterion for acceptance.

Articles may be expositions of old results or presentations of new ones. They may concern all of mathematics or one small area, a broad development or a single application, historical reminiscences or one important event. While some articles may contain the author's new research, the novelty of material and generality of the results is far less important than the clarity of exposition and general interest. Discussing one illuminating case of a well known result is far better than providing all the details of an obscure but new proposition. Articles in the *Monthly* are supposed to inform and to entertain; they are meant to be read rather than archived.

Notes are short and possibly informal articles. A note may concern a clever new proof of an old theorem, a novel way to present tired material, or a lively discussion of a philosophical (but still mathematical) issue. Also, any topic is suitable, so long as it is related to mathematics. Because a note is short, the first few sentences are the most important part: They should explain the purpose and invite the reader in. Photographs or diagrams often will attract the reader's attention.

All articles and notes should be sent to the editor:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Please send 3 copies, typewritten on only one side of the paper. Illustrations should be carefully drawn on separate sheets of paper in black ink; the original should be without lettering and two copies should have appropriate captions and lettering indicated.

Proposed problems or solutions should be sent to:

RICHARD BUMBY,
P.O. Box 10971
New Brunswick, NJ 08906-0971.

Please send 2 copies of all material, typewritten if possible.

Letters to the Editor, both for publication and for private reading, should be sent to the Editor at the address given above. Comments, including criticisms, are welcome, as are all suggestions for making the *Monthly* a lively, entertaining, and informative journal.

COVER PHOTO: A not-so-random dot stereogram showing a hemisphere in center. It works best to hold the picture about 18" from your eyes.

EDITOR:

JOHN H. EWING

ASSOCIATE EDITORS:

PETER BORWEIN	FRED KOCHMAN
RICHARD BUMBY	CATHERINE MCGEOCH
DENNIS DETURCK	RICHARD NOWAKOWSKI
UNDERWOOD DUDLEY	ARNOLD OSTEBEE
JOHN DUNCAN	LEE RUBEL
JOAN FERRINI-MUNDY	ABE SHENITZER
JOSEPH GALLIAN	LYNN STEEN
STEVEN GALOVICH	STAN WAGON
RICHARD GUY	DOUGLAS WEST
DARRELL HAILE	HERBERT WILF
PAUL HALMOS	SANDY ZABELL
JOAN HUTCHINSON	PAUL ZORN

EDITORIAL ASSISTANT:

MISTY CUMMINGS

STAFF ARTIST:

MIKE CAGLE

Reprint permission:

MARCIA P. SWARD, Executive Director

Advertising Correspondence:

Ms. ELAINE PEDREIRA, Advertising Manager

Subscription correspondence, change of address, and other inquiries:

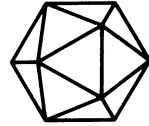
Membership / Subscriptions Department

All at the address:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036.

Microfilm Editions: University Microfilms International, Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT. Copyrighted by the Mathematical Association of America (Incorporated), 1994, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source. Second class postage paid at Washington, DC, and additional mailing offices. **Postmaster:** Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.



Contents

ARTICLES

**Behind the Scenes of a Random Dot Stereogram / MARIA S. TERRELL
and ROBERT E. TERRELL 715**

**The Fifty-Fourth William Lowell Putnam Mathematical Competition /
LEONARD F. KLOSINSKI, GERALD L. ALEXANDERSON,
and LOREN C. LARSON 725**

Literacy in the Language of Mathematics / JAMES O. BULLOCK 735

**Fractional and Trigonometric Expressions for Matrices /
GORO SHIMURA 744**

**Noether Lasker Primary Decomposition Revisited /
BARBARA L. OSOFSKY 759**

**Elementary Infinite Sources of Non-Unique Factorization Rings / S. STEIN
and S. SZABÓ 769**

***Apropos* Two Notes on Notation / ANTAL E. FEKETE 771**

FEATURES

COMMENTS 714

NOTES

**The Coin Exchange Problem for Arithmetic Progressions /
AMITABHA TRIPATHI 779**

Congruence of Triangles / LEONARD GILLMAN 782

**A Short Elementary Proof of the Mohr-Mascheroni Theorem /
NORBERT HUNGERBÜHLER 784**

UNSOLVED PROBLEMS

**Which Triangles Are Plane Sections of Regular Tetrahedra? /
FOLKE ERIKSSON 788**

THE AUTHORS 790

PROBLEMS AND SOLUTIONS 792

REVIEWS

***Brouwer's Intuitionism.* By Walter P. van Stigt / C. SMORYŃSKI 799**

TELEGRAPHIC REVIEWS 803

Comment

See Response on page 809

The recent report *Recognition and rewards in the mathematical sciences* is a strange and pedantic document. It comes from a committee of the JPBM (The Joint Policy Board for Mathematics). It is gussied up in slick paper and shaded bar graphs. (Evidently, readers were not expected to read numbers or percents.) It ignores the long standing standards of the American mathematical community. These standards read simply: Do good mathematics. This has meant good teaching (tell the students when you use the Mean Value Theorem), good research (proofs that are complete), good guidance (help troubled students to understand math), and good judgment, both of colleagues and of research results.

The report ignores these powerful traditions and instead follows insignificant ideas from Ernie Boyer. Boyer wants to "reconsider" scholarship, so this report has an elaborate and pedantic discussion of a definition (*sic*) of "mathematical scholarship". Each department is urged to provide its own definition, in keeping with its so-called "mission" (a presently popular buzz-word).

The JPBM committee conducted a survey. The report does not tell us what questions were asked; since the exact form of the questions is vital, this omission is irresponsible. For example (p. 4) "with regard to research, we asked ... how it was evaluated" — not whether it was good mathematics or accurate evaluation.

Questions were asked separately of chairs (*sic*) and faculty. One question may have been "How important is classroom teaching in determining merit salary raises" (p. 10), but figure 13 has "percent responding that salary reflects differences between excellent and average teaching". What question did they really ask? The committee professes surprise at "The consistent difference in the perception of the chairs and the faculty. Figure 13 shows that a much higher percentage of chairs than faculty believed differences in teaching effectiveness were truly reflected in salary. Why the surprise? Chair people often help set salary, and faculty can be jealous. And did the question really use the word "Truly"?

The committee discussion of mathematical research (p. 28) is wholly inadequate; no Fermat, no quantum fields, no motives. One original member of the committee resigned.

The committee claims (p. 7) a "consensus" that "faculty members who change emphasis during their career (e.g., from research to education)... should be allowed to do so without being penalized by the rewards system." This is hopelessly vague. What education; a poor text book? Was it good research? What type of university? Is there really a "rewards System"?

The report favors, with little or no evidence, the institution of "periodic review of faculty members". Does the background of the report involve some "Hidden Agenda"? There is reference to a 1990 book by Boyer and to publications by Robert M. Diamond. He is Assistant Vice-Chancellor at Syracuse University, and he has organized in the Adirondacks an interdisciplinary meeting on "faculty rewards" (*cf* Chronicle of Higher Ed., May 11, 1994). I recently heard Dr. Diamond talk at a symposium. He opposed tenure and "wished to define priorities" and considered what faculty members are doing in their field. Do we wish Chancellors, vice or not, to direct our research?

This JPBM report is a striking example of the current practice of selected pundits telling the rest of us what we should do. It was prepared by a small and sometimes secretive group (as in an incomplete presentation at Cincinnati). It is an assault on the standards of our community.

Saunders Mac Lane

Response

See Comment on page 714

The commentary of Saunders Mac Lane on page 714 raises a number of points related to the contents of the Recognition and Rewards report and, as we read it, to the appropriateness of engaging in any discussion about how our community has been able to encourage its members to engage in the broad spectrum of professional duties that are part of the responsibilities of a department or the university in which it is situated.

Mac Lane has raised some questions as to the specifics of the survey and its interpretation. The full data and survey instruments were too extensive to include in the report but, as stated on page four of the report, the American Mathematical Society will provide them on request. Indeed, we would be pleased to have them examined and discussed by many members of the community.

Why the report and why now? Colleges and universities today teach the daughters and sons of a mass society, a far different population than was the case when most of this structure was put in place. As educators we are being called upon to do a better job. These calls are coming from students, parents, legislators, from academic leaders such as Don Kennedy and from members (most of them academics) of a recent President's Council of Advisors on Science and Technology (cf. *Renewing the Promise: Research Intensive Universities and the Nation*). As researchers we are being called upon to address issues of national importance and perhaps, as Roland Schmitt and the members of PCAST cited above suggest, scientists can and should take that responsibility, and that it is possible to so whilst preserving our core values. It is against just such a background that the JPBM charged a committee to initiate a dialogue encompassing "differing views about the need for change and about how well the system works now".

The report finds that the work of professors involves much more than research and that in all too many cases the selection, evaluation and promotion of faculty relies too much on the one dimensional criterion of research, and, in so doing, fails to recognize or encourage other contributions necessary in order to meet the mission of the department or the university in which it is situated. Mac Lane's apparent scorn for departmental "mission" misses the point. The Committee clearly recognizes the diversity of institutions and attendant faculty responsibilities. As is clearly stated, the report is mainly addressed to individual departments encouraging the faculty to discuss among themselves what it is they are about, what their goals are and what they want to achieve. It is simply false that the report suggests or recommends that faculty be rewarded for work that lacks quality. Indeed this is a principal reason for seeking a statement of scholarship and for the attention paid by the Committee to evaluation. As to the work of Boyer and Diamond, it is accepted practice to place work in context by referencing previous work.

Rather than "an assault on the standards of the community" the report is a call to engage in discussion of issues that are upon us. It is only through such discussion that we can ensure the health and viability of our community.

Richard H. Herman,
Chair JPBM

Calvin C. Moore, *Chair, JPBM Committee*
on Professional Recognition and Rewards

Behind the Scenes of a Random Dot Stereogram

Maria S. Terrell and Robert E. Terrell

What are all those people staring at? What do they hope to see in those finely textured posters, beyond the almost random display of little flowers or paint splotches? “Look deeper into the picture.” “Look beyond the surface.” “Cross your eyes.” “Blur your vision.” Not since Rubic’s Cube has a puzzle or game engaged and fascinated the public as have the currently faddish 3D illusions on display in bookstores and shopping malls across the country. For many the exercise ends in frustration, seeing nothing beyond the surface images. Could it be a case of mass hypnosis or “The Emperor’s New Clothes”? For those who see the three dimensional images pop into view, there is surprise, pleasure, and wonder. Having had experiences of both types, we sympathize with those who have decided they have better things to do than stare off into space. But in fact knowing how the pictures are created, understanding the elementary geometry and optics of stereographic vision, makes it much easier to see the images.

We begin with the simplest example. Stare at the two dots below. If you relax your eyes you will probably see four dots. Try to merge the two middle images so that you see three dots. Once you’ve done that, you’ve seen your first, not so random, random dot stereogram. If after a few minutes you find that you cannot make yourself see three dots, you may be one of the 2% of the population which is stereoblind.



Random dot stereograms are the offspring of an unlikely marriage of aerial reconnaissance and brain research. Late in the 1950’s the experiences of psychologist and former radar engineer, Bela Julesz, led him to challenge prevailing theories of depth perception [1]. Julesz knew that when aerial photographs of camouflaged areas, taken from two different positions, were viewed through a stereoscope, tanks and equipment appeared to pop out of the almost random patterns. This convinced him that depth perception could occur without monocular visual cues. It was a revolutionary idea. He proved that he was right by devising the random dot stereogram, a pair of random arrays one for each eye, which when viewed simultaneously created a three dimensional image. The pair of random dot arrays below are examples of Julesz’s ground breaking tool. Without special



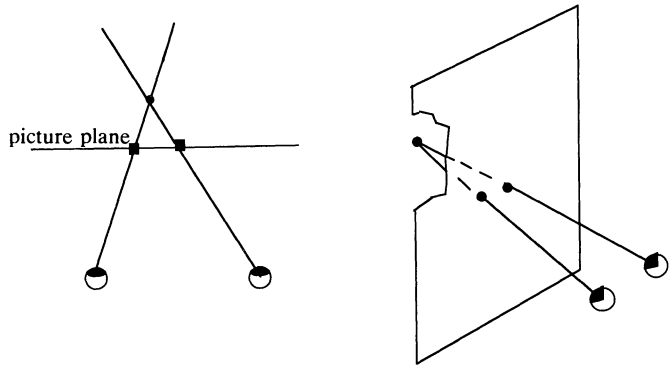
Random dot stereogram from Julesz, 1971.

equipment it is fairly difficult to see the 3D image of a diamond hovering above a background plane. If you'd like to try, relax your eyes and make three arrays from two like you did with the dots.

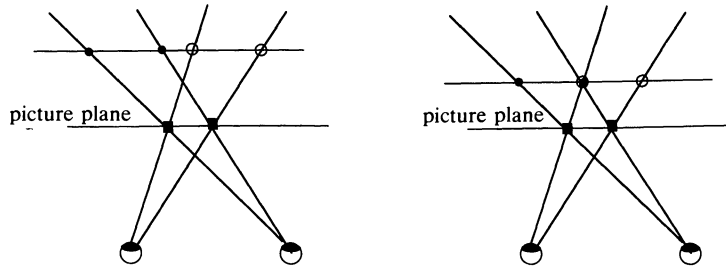
In addition to the separate stereo images Julesz [2] combined the left and right eye views by coloring one green and the other red. When viewed through glasses with one red and one green lense the 3D images popped out of the array. It is hard to overstate the importance of random dot stereograms as a tool to study one aspect of depth perception without having to sort out and account for multiple effects of various visual cues.

In 1983 C. W. Tyler [3] described a way of combining the left and right eye information into one picture called the *auto* random dot stereograms, which did not require the use of special equipment. Rather than superimposing the left and right eye views, as Julesz did, Tyler found a way to make one picture which the brain can interpret as a 3D image. Although known by various commercial names in the marketplace, we refer to these pictures simply as *stereograms*. We have dropped the term “random” because the dots do not have to be random. Also the dots do not have to be dots, but can be small circles or other figures.

Before discussing how the more interesting stereograms are made, consider the case of creating a stereogram of a single point. Think of the drawing below as a view from above of the eyes, the two dots on the picture, and the single point behind the picture.

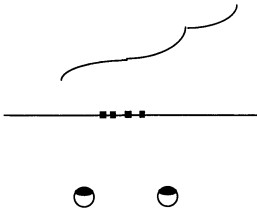


Since each of the dots is seen by both eyes, there are additional lines of sight through them. Each eye has a retinal image of the two dots which the brain can interpret as three or four dots at various depths behind the picture plane. While there are many places at which the two dots look like four, there's only one at which they look like three. Merging the images of the two dots, completely determines the position of the perceived dot in the space behind the picture.

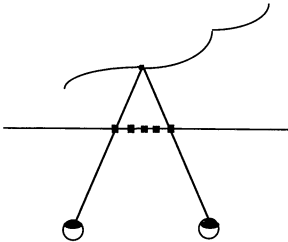


There is another point of intersection, but it determines an image which appears to be in front of the stereogram. Since most of the commercial stereograms seem to be designed to depict the image behind the plane, we will concentrate on that image. However, results like those we derive for the image behind the plane can be applied to the image in front as well.

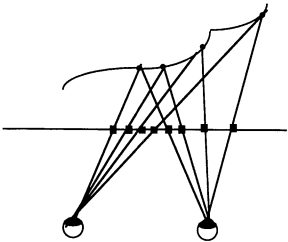
The next sequence of figures gives an example of how to construct one row of a stereogram which will represent the curve shown, starting from an interval of dots.



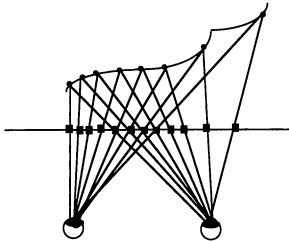
Draw a line from the left eye through one of the dots in the interval and mark the point at which that line intersects the curve. Next draw the line of sight from this point on the curve to the right eye. Put a new dot at the point at which this line intersects the picture plane.



Repeat this process with the other dots in the original interval. The resulting eight dots in the stereogram, represent four points on the curve.

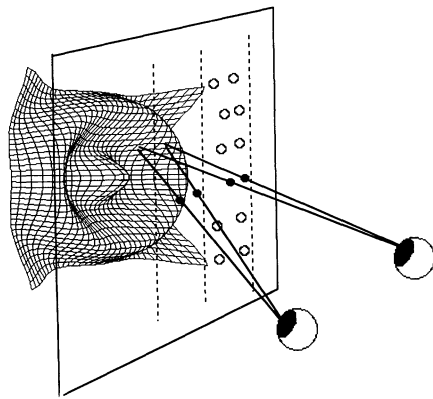


Since the original four dots also represent four points on the curve as seen by the right eye, they determine four new left eye images.

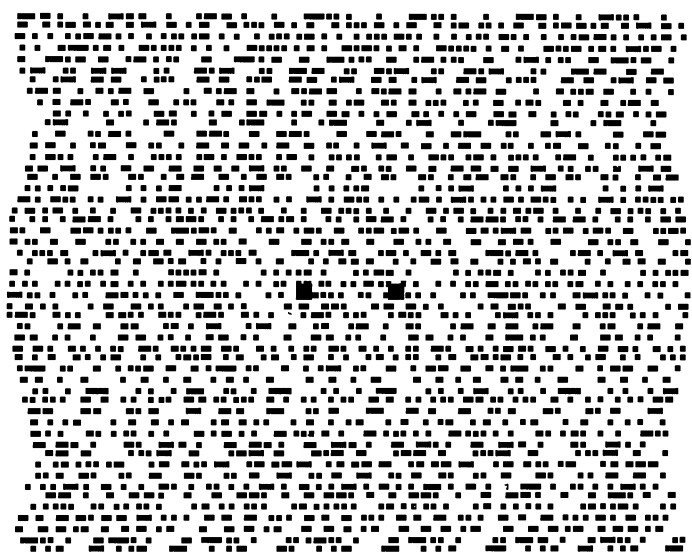


If the curve were longer, we would continue working to the left and right making more intervals of dots. Most stereograms have at least five such intervals in each row. The outermost intervals represent points seen by only one eye.

To create a two dimensional representation of a three dimensional scene, we stack rows of dots which were generated as above. If the intervals of initial dots are all from the center, they form a vertical strip, from which the rest of the display is generated. The initial dots do not have to lie in a vertical strip, but it is a convenience.

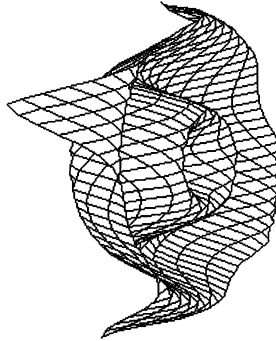


Now that you have some idea of the principles involved in generating these pictures, try seeing the “Mexican Hat”. We have included a set of guide dots which you may, or may not find useful. To use them make the two dots look like three. This may help you fixate at a point behind the picture and is more precise than the commercial advise to look at your reflection. When you begin to sense some depth, relax and look for a surface similar to the one in the figure above. A word of caution—don’t expect to see anything if you rotate the stereogram. The underlying principle is that your eyes are separated horizontally and that your brain is quite accustomed to fusing horizontally disparate retinal images.

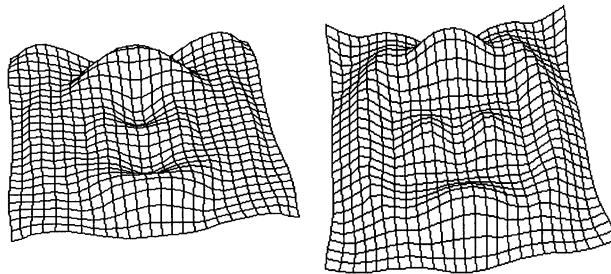


Mexican Hat

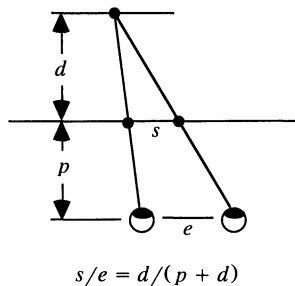
Sometimes when you look at a stereogram, a scene comes into view which is not at all what you thought it should be. This does not mean there is something wrong with your eyes! You may be seeing the front image that we mentioned earlier. It is similar to a reflection of the one behind the plane. To see it, hold a pencil a few inches in front of the picture, so that when you're looking at the stereogram, two somewhat transparent pencils have their tips directly under the two guide dots. Now fixate on the tip of the pencil, forcing its two images into one. The two dots will now look like three, with one directly over the tip of the pencil. As you fix your attention on the pencil point, you'll notice that a 3D image is beginning to emerge nearby. It should look roughly like the image below. Once you've got it, you can try poking the pencil through it!



In addition to the primary images above, it is possible to perceive something like the shapes shown in the next figures at depths farther behind (or reversed and in front) of the picture plane.



The perception of these deeper images is easier to describe after looking at the perception of a single image more carefully. In the plane through the eyes and a point in the scene, we use similar triangles to find a relation among the distances e = eye spacing, s = dot spacing, p = distance from your eyes to the line through the dots, and d = distance from the point to the line.

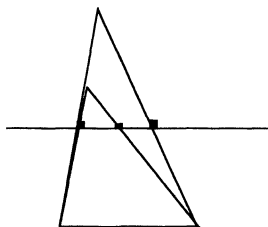


A number of consequences can be predicted from this geometric model. First, the perceived depth of a point in the 3D scene is determined by the distance between its image dots. As a result, stereograms of shallow scenes have fairly strong repeated patterns. Conversely, perfectly regular repeated spacing of figures can cause the illusion of a recessed plane, called the wallpaper effect.

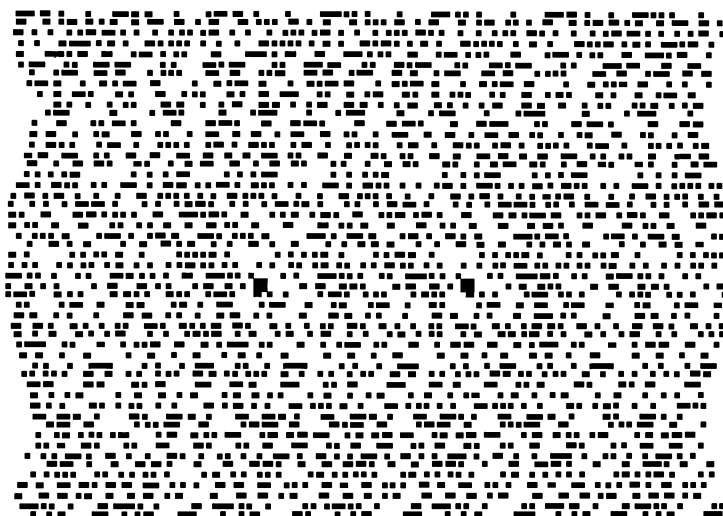
Moving the stereogram closer to or farther from your eyes may help you see the image, because it allows you to fixate at different depths. Since this does not change s or e , similar triangles predict that d increases if p does. Once you have seen an image, try slowly moving the stereogram back and forth to experience changes in the depth of the perceived image.

This simple relationship should also be considered when choosing the strip width for a particular scene. If the strip is too narrow, relative to the depth of the scene, it will lead to gaps in the stereogram, vertical spaces with no dots at all. If it is too wide, it may lead to new dots falling in the initial interval, perhaps making solid black regions.

Under what circumstances might you see a completely different second image behind the picture? Suppose the strip width is less than half the distance between your eyes. If the stereogram was generated by producing eight strips of that width, then there are four strips of about twice that width.

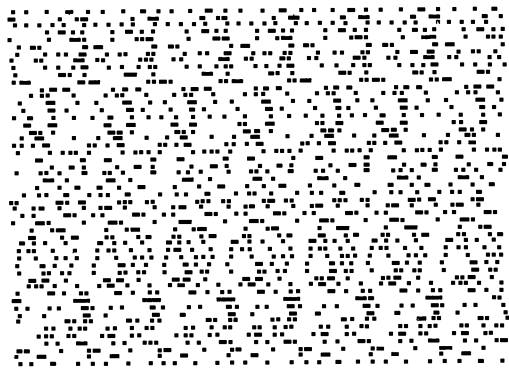


To see a third image it is necessary that $3g < e$, and so forth. If you have had success so far, you might try to see other images in the Mexican Hat stereogram to which we have added double wide guide dots.

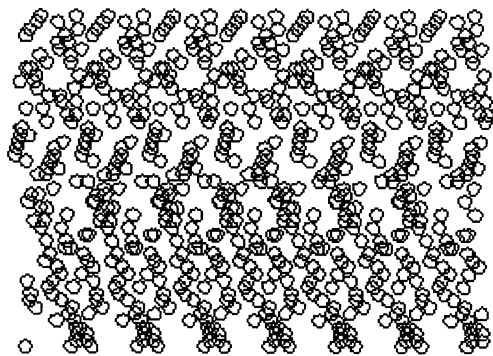


If you are looking at a stereogram with wide dot spacing and you can see only one or two images in it, take it to a copy machine and reduce the size. Then you

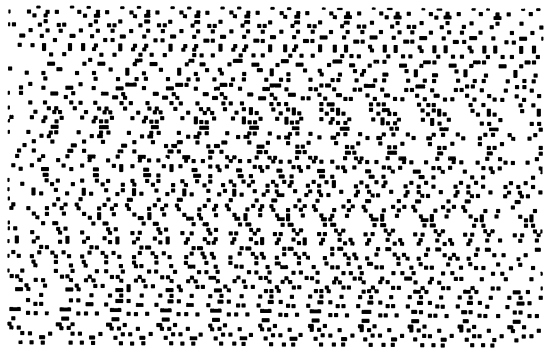
might see a third and fourth image in it. Try it. We have never seen more than five deep in any one picture. Now that you are experienced, try these stereograms, which we have labeled to suggest their 3D images. If you find that you need guide dots, find a repeated pattern within one row and color any two of the corresponding dots. We have included pictures of the underlying surfaces and information about their sets of initial dots at the end of this article.



Eggcrate

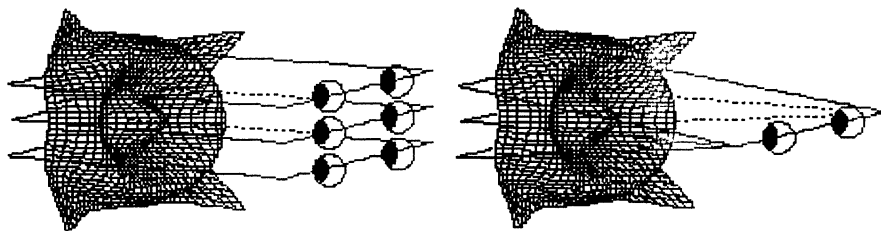


Pocket



Bug

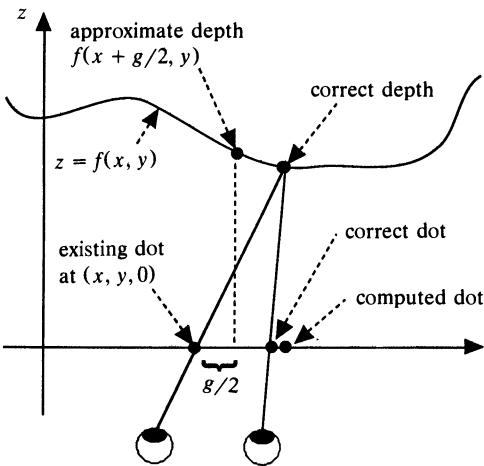
The algorithm uses intersections of the scene with horizontal, rather than sloping planes through the eyes. This is one source of error.



At the planar level we approximate the spacing for the new dot by

$$s = ef(x + g/2, y)/(p + f(x + g/2, y))$$

where g is the width of the generating strip, e is the distance between the eyes, and p is the distance from the eyes to the picture plane, (x, y) are coordinates in the picture plane, and the graph of f is the scene depicted.



Our stereograms were computed using an algorithm similar to the one described in [4]. It uses two approximations which avoid the difficulty of computing intersections of lines and surfaces. *See box.* These approximations do cause some distortion, but as long as the objects are near the center of the visual field, it is usually not too noticeable. In the example below two hemispheres appear to sag toward the center, when viewed from a point fairly close to the center of the picture. When viewed from points more directly over the hemispheres, the distortion is reduced.

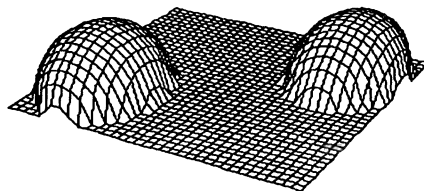


Hemispheres

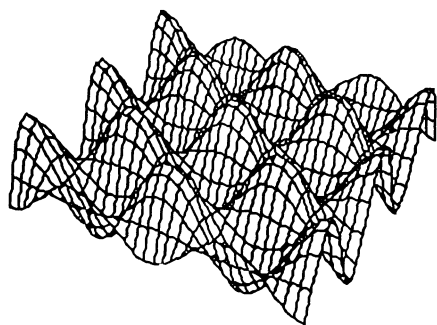
There are many questions to explore and images to create with stereograms. Psychologists continue to use both stereo pairs and single stereograms for research in depth perception. For us the existence of the deeper images and their dependence on both the primary surface and the dot spacing algorithm, continues to be a source of surprise and curiosity.

REFERENCES

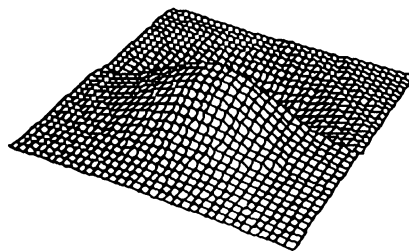
1. B. Julesz, *Foundations of Cyclopean Perception*. The University of Chicago Press, Chicago (1971).
2. B. Julesz, Stereoscopic Vision, *Vision Research* 26 (1986) 1601–1611.
3. C. Tyler, Sensory Processing of Binocular Disparity, in *Vergence Eye Movements: Basic and Clinical Aspects*. Butterworth, Boston (1983) 199–295.
4. D. Bar-Natan, Random-Dot stereograms, *Mathematica Journal* 1 (1991), 69–71.



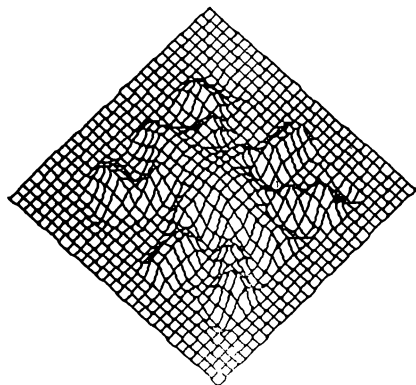
Hemispheres



Eggcrate



Pocket



Bug

Notes. The hemispheres, pocket, and eggcrate were constructed from a random vertical center strip. The bug was generated from a random scattering of initial points over the whole area of the picture. The pocket was rendered using a small circle instead of a dot at each computed position.

*Department of Mathematics
Cornell University
Ithaca, NY 14853
maria@math.cornell.edu*

The Fifty-Fourth William Lowell Putnam Mathematical Competition

Leonard F. Klosinski, Gerald L. Alexanderson,
and Loren C. Larson

The following results of the fifty-fourth William Lowell Putnam Mathematical Competition, held on December 4, 1993, have been determined in accordance with the governing regulations. This annual contest is supported by the William Lowell Putnam Prize Fund for the Promotion of Scholarship, left by Mrs. Putnam in memory of her husband, and is held under the auspices of the Mathematical Association of America.

The first prize, \$7,500, was awarded to the Department of Mathematics of Duke University. The members of the winning team were: Andrew O. Dittmer, Craig B. Gentry, and Jeffrey M. Vanderkam; each was awarded a prize of \$500.

The second prize, \$5,000, was awarded to the Department of Mathematics of Harvard University. The members of the winning team were Kiran S. Kedlaya, Serban M. Nacu, and Royce Y. Peng; each was awarded a prize of \$400.

The third prize, \$3,000, was awarded to the Department of Mathematics of Miami University. The members of the winning team were John D. Davenport, Jason A. Howald, and Matthew D. Wolf; each was awarded a prize of \$300.

The fourth prize, \$2,000, was awarded to the Department of Mathematics of the Massachusetts Institute of Technology. The members of the winning team were Henry L. Cohn, Alexandru D. Ionescu, and Andrew Przeworski; each was awarded a prize of \$200.

The fifth prize, \$1,000, was awarded to the Department of Mathematics of the University of Michigan, Ann Arbor. The members of the winning team were Philip L. Beineke, Brian D. Ewald, and Soundararajan Kannan.

The six highest ranking individual contestants, in alphabetical order, were Craig B. Gentry, Duke University; J. P. Grossman, University of Toronto; Wei-Hwa Huang, California Institute of Technology; Kiran S. Kedlaya, Harvard University; Adam M. Logan, Princeton University; and Lenhard L. Ng, Harvard University. Each of these was designated a Putnam Fellow by the Mathematical Association of America and awarded a prize of \$1,000, by the Putnam Prize Fund.

The next three highest ranking contestants, in alphabetical order, were Michail Sunitzky, Princeton University; Dylan P. Thurston, Harvard University; and Jade P. Vinson, Washington University, St. Louis; each was awarded a prize of \$500.

The next six highest ranking contestants, in alphabetical order, were Mikhail Kogan, New York University; Akira Negi, University of North Carolina, Chapel Hill; Joel E. Rosenberg, Princeton University; David L. Savitt, University of British Columbia; Jeffrey D. Wall, Princeton University; and Thomas A. Weston, Massachusetts Institute of Technology; each was awarded a prize of \$250.

The next thirteen highest ranking contestants, in alphabetical order, were Manjul Bhargava, Harvard University; John D. Davenport, Miami University; Andrew O. Dittmer, Duke University; Sergey V. Levin, Harvard University; Douglas A. Levy, University of Pennsylvania; William R. Mann, Princeton University; Adam W. Meyerson, Massachusetts Institute of Technology; Curtis Z. Mitchell, Carleton College; Serban M. Nacu, Harvard University; An T. Nguyen, University of Texas, Austin; Byron M. Shock, Albertson College of Idaho; Ka-Ping R. Yee, University of Waterloo; and Douglas J. Zare, New College of the University of South Florida; each was awarded a prize of \$100.

The following teams, named in alphabetical order, received honorable mention: Cornell University, with team members Robert D. Kleinberg, Mark Krosky, and Tong Zhang; New York University, with team members Igor Berger, Yevgeniy Dodis, and Mikhail Kogan; Princeton University, with team members Ze Y. Chen, Adam M. Logan, and William R. Mann; University of Toronto, with team members J. P. Grossman, Edwin N. Sato, and Hugh Thomas; and the University of Waterloo, with team members Daniel R. L. Brown, Ian A. Goldberg, and Peter L. Milley.

Honorable mention was achieved by the following twenty-nine individuals named in alphabetical order: Jared E. Anderson, University of Victoria; Jonathan E. Atkins, Rose-Hulman Institute of Technology; Henry L. Cohn, Massachusetts Institute of Technology; Ilya A. Entin, Massachusetts Institute of Technology; Brian D. Ewald, University of Michigan, Ann Arbor; Kevin E. Foltz, Rice University; David Friedman, Massachusetts Institute of Technology; Ian A. Goldberg, University of Waterloo; H. Tracy Hall, Brigham Young University; John D. Harrington, University of Idaho; Simeon J. Hellerman, Brown University; Timothy J. Hollebeek, Calvin College; Alexandru D. Ionescu, Massachusetts Institute of Technology; Daniel C. Isaksen, University of California, Berkeley; Soundararajan Kannan, University of Michigan, Ann Arbor; Robert D. Kleinberg, Cornell University; Botand Kőszegi, Harvard University; Josh L. Levenberg, Reed College; Jie J. Lou, University of Waterloo; Idris D. Mercer, University of Victoria; Frosti Petursson, University of Pennsylvania; Anand J. Reddy, University of California, Berkeley; Edwin N. Sato, University of Toronto; Jason R. Schweinsberg, Williams College; Mark A. Van Raamsdonk, University of British Columbia; Jeffrey M. Vanderkam, Duke University; Wayne A. Whitney, Harvard University; Tong Zhang, Cornell University; and Zhaohui Zhang, Yale University.

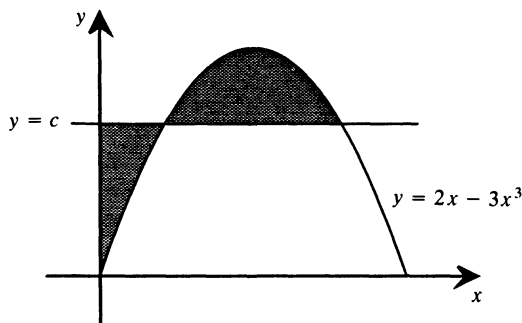
The other individuals who achieved ranks among the top 101, in alphabetical order of their schools, were: University of Arizona, Randy W. Ho; Brigham Young University, John Wesley Robertson; California Institute of Technology, Julian C. Jamison; University of California, San Diego, David R. Wasserman; University of Chicago, Dean W. Jens; Colorado State University, Matthew K. Kahle; Cornell University, Demetrio A. Muñoz; Gordon College, Hai Shao; Harvard University, Swaine L. Chen, Hank S. Chien, Adam Kalai, Joseph D. Kanapka, Joshua N. Newman, David S. Patterson, Royce Y. Peng, Jonathan E. Tannenhausser; Massachusetts Institute of Technology, Andrew Przeworski, Dmitriy A. Rogozhnikov, Jason M. Sachs, Ritesh A. Shah; McGill University, Rajesh J. Pereira; Miami University, Jason A. Howald, Matthew D. Wolf; University of Michigan, Ann Arbor, Philip L. Beineke; University of Minnesota, Minneapolis, Matthew P. Kelly; New York University, Yevgeniy Dodis; North Carolina State University, Stephen A. London; University of Northern Colorado, John C. Petherick; Princeton University, Steven S. Gubser, Mark W. Lucianovic, Thomas J. Weisswange; University of Saskatchewan, Trevor N. Green; Simon Fraser University, Erick B. Wong; Stanford University, Svetlozar E. Nestorov; Swarthmore College, Mark D. Kernighan; Texas Tech University, Mikhail V. Shubov; Vassar College, Andrew F. Rizzo; Washington University, St. Louis, Benjamin B. Gum, Edward D. Hanson, Philip X. Wu; University of Waterloo, Daniel R. L. Brown, Eli Lapell; University of Wisconsin, Madison, Brent E. Halsey; and University of Wisconsin, Parkside, Sergey M. Ioffe.

There were 2356 individual contestants from 402 colleges and universities in Canada and the United States in the competition of December 4, 1993. Teams were entered by 291 institutions.

The Questions Committee for the fifty-fourth competition consisted of George T. Gilbert, Texas Christian University (Chair); Fan Chung, Bellcore; and Eugene Luks, University of Oregon; they composed the problems listed below and were most prominent among those suggesting solutions.

PROBLEMS

Problem A-1. The horizontal line $y = c$ intersects the curve $y = 2x - 3x^3$ in the first quadrant as in the figure. Find c so that the areas of the two shaded regions are equal.



Problem A-2. Let $(x_n)_{n \geq 0}$ be a sequence of nonzero real numbers such that

$$x_n^2 - x_{n-1}x_{n+1} = 1, \quad \text{for } n = 1, 2, 3, \dots.$$

Prove there exists a real number a such that $x_{n+1} = ax_n - x_{n-1}$ for all $n \geq 1$.

Problem A-3. Let \mathcal{P}_n be the set of subsets of $\{1, 2, \dots, n\}$. Let $c(n, m)$ be the number of functions $f: \mathcal{P}_n \rightarrow \{1, 2, \dots, m\}$ such that $f(A \cap B) = \min\{f(A), f(B)\}$. Prove that

$$c(n, m) = \sum_{j=1}^m j^n.$$

Problem A-4. Let x_1, x_2, \dots, x_{19} be positive integers each of which is less than or equal to 93. Let y_1, y_2, \dots, y_{93} be positive integers each of which is less than or equal to 19. Prove that there exists a (nonempty) sum of some x_i 's equal to a sum of some y_j 's.

Problem A-5. Show that

$$\int_{-100}^{-10} \left(\frac{x^2 - x}{x^3 - 3x + 1} \right)^2 dx + \int_{\frac{1}{101}}^{\frac{1}{11}} \left(\frac{x^2 - x}{x^3 - 3x + 1} \right)^2 dx + \int_{\frac{101}{100}}^{\frac{11}{10}} \left(\frac{x^2 - x}{x^3 - 3x + 1} \right)^2 dx$$

is a rational number.

Problem A-6. The infinite sequence of 2's and 3's

2, 3, 3, 2, 3, 3, 3, 2, 3, 3, 3, 2, 3, 3, 2, 3, 3, 3, 2, 3, 3, 3, 2, 3, 3, 3, 2, 3, 3, 2, 3, 3, 3, 2, ...

has the property that, if one forms a second sequence that records the number of 3's between successive 2's, the result is identical to the given sequence. Show that there exists a real number r such that, for any n , the n th term of the sequence is 2

if and only if $n = 1 + \lfloor rm \rfloor$ for some nonnegative integer m . (Note: $\lfloor x \rfloor$ denotes the largest integer less than or equal to x .)

Problem B-1. Find the smallest positive integer n such that for every integer m , with $0 < m < 1993$, there exists an integer k for which

$$\frac{m}{1993} < \frac{k}{n} < \frac{m+1}{1994}.$$

Problem B-2. Consider the following game played with a deck of $2n$ cards numbered from 1 to $2n$. The deck is randomly shuffled and n cards are dealt to each of two players, A and B . Beginning with A , the players take turns discarding one of their remaining cards and announcing its number. The game ends as soon as the sum of the numbers on the discarded cards is divisible by $2n + 1$. The last person to discard wins the game. Assuming optimal strategy by both A and B , what is the probability that A wins?

Problem B-3. Two real numbers x and y are chosen at random in the interval $(0,1)$ with respect to the uniform distribution. What is the probability that the closest integer to x/y is even? Express the answer in the form $r + s\pi$, where r and s are rational numbers.

Problem B-4. The function $K(x,y)$ is positive and continuous for $0 \leq x \leq 1, 0 \leq y \leq 1$, and the functions $f(x)$ and $g(x)$ are positive and continuous for $0 \leq x \leq 1$. Suppose that for all $x, 0 \leq x \leq 1$,

$$\int_0^1 f(y)K(x,y) dy = g(x) \quad \text{and} \quad \int_0^1 g(y)K(x,y) dy = f(x).$$

Show that $f(x) = g(x)$ for $0 \leq x \leq 1$.

Problem B-5. Show there do not exist four points in the Euclidean plane such that the pairwise distances between the points are all odd integers.

Problem B-6. Let S be a set of three, not necessarily distinct, positive integers. Show that one can transform S into a set containing 0 by a finite number of applications of the following rule: Select two of the three integers, say x and y , where $x \leq y$, and replace them with $2x$ and $y - x$.

SOLUTIONS

In the 12-tuples $(n_{10}, n_9, n_8, n_7, n_6, n_5, n_4, n_3, n_2, n_1, n_0, n_{-1})$ following each problem number below, n_i for $10 \geq i \geq 0$ is the number of students among the top 207 contestants achieving i points for the problem and n_{-1} is the number of those not submitting solutions.

A-1 (185, 2, 0, 0, 0, 0, 0, 0, 1, 0, 14, 5)

Solution. The value of c is $4/9$.

Let (b, c) denote the second intersection point. We wish to find c so that

$$\int_0^b (c - (2x - 3x^3)) dx = 0.$$

This leads to $cb - b^2 + (3/4)b^4 = 0$. After substituting $c = 2b - 3b^3$ and solving, we find that $b = 2/3$ and the result follows.

A-2 (146, 21, 6, 0, 0, 0, 0, 8, 1, 17, 8)

Solution 1. It is equivalent to show that

$$\frac{x_{n+1} + x_{n-1}}{x_n}$$

is independent of n . This follows (by induction) from

$$\begin{aligned} \frac{x_{n+2} + x_n}{x_{n+1}} - \frac{x_{n+1} + x_{n-1}}{x_n} &= \frac{(x_n x_{n+2} + x_n^2) - (x_{n+1}^2 + x_n x_{n+1})}{x_n x_{n+1}} \\ &= \frac{-(x_{n+1}^2 - x_n x_{n+2}) + (x_n^2 - x_{n-1} x_{n+1})}{x_n x_{n+1}} \\ &= \frac{-1 + 1}{x_n x_{n+1}} = 0. \end{aligned}$$

Solution 2. For all n ,

$$\begin{aligned} \det \begin{pmatrix} x_{n-1} + x_{n+1} & x_n + x_{n+2} \\ x_n & x_{n+1} \end{pmatrix} &= \det \begin{pmatrix} x_{n-1} & x_n \\ x_n & x_{n+1} \end{pmatrix} + \det \begin{pmatrix} x_{n+1} & x_{n+2} \\ x_n & x_{n+1} \end{pmatrix} \\ &= -1 + 1 = 0. \end{aligned}$$

Thus, $(x_{n-1} + x_{n+1}, x_n + x_{n+2}) = c_n(x_n, x_{n+1})$ for some scalar c_n . Substituting $n - 1$ for n and then comparing the coordinate expressions, we see that $c_n = c_{n-1}$ (using $x_n \neq 0$).

Comment. In a similar manner, one can prove that if $(x_n)_{n \geq 0}$ is a sequence of nonzero real numbers such that

$$\det \begin{pmatrix} x_n & x_{n+1} & \cdots & x_{n+k} \\ x_{n+1} & x_{n+2} & \cdots & x_{n+k+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n+k} & x_{n+k+1} & \cdots & x_{n+2k} \end{pmatrix} = cr^n \quad \text{for } n = 1, 2, 3, \dots,$$

then there exist real numbers a_1, \dots, a_k such that

$$x_{n+k+1} = a_1 x_{n+k} + a_2 x_{n+k-1} + \cdots + a_k x_{n+1} + (-1)^k r x_n.$$

A-3 (2, 11, 24, 0, 0, 0, 0, 27, 8, 54, 81)

Solution. Let $S = \{1, 2, \dots, n\}$, and suppose that f is such a function and that $f(S) = j$. Then $f(A \cap B) = \min\{f(A), f(B)\}$ implies that the values of $f(S - \{i\})$ determine $f(A)$ for all $A \subset S$, since $f(A) = \min_{i \notin A} \{f(S - \{i\})\}$ for $A \subset S$. Conversely, arbitrary choices of $f(S - \{i\}) \leq j$ leads to such a function. Because there are j independent choices for each $f(S - \{i\})$, there are j^n functions with $f(S) = j$. Summing over the possible values of $f(S)$ yields $c(n, m) = \sum_{j=1}^m j^n$.

A-4 (3, 2, 0, 0, 0, 0, 0, 0, 0, 44, 158)

Solution. For reasons of symmetry, let us replace 19 and 93 by m and n respectively, in the problem statement. Without loss of generality, $\sum_{i=1}^m x_i \geq \sum_{j=1}^n y_j$.

Then, for $0 \leq k \leq n$, there exists $f(k)$, $0 \leq f(k) \leq m$, such that

$$\sum_{i=1}^{f(k)} x_i \leq \sum_{j=1}^k y_j < \sum_{i=1}^{f(k)+1} x_i.$$

Let $g(k) = \sum_{j=1}^k y_j - \sum_{i=1}^{f(k)} x_i$. Then, for $0 \leq k \leq n$, $0 \leq g(k) < x_{f(k)+1} \leq n$. If $g(k) = 0$ for some k , we are done. Otherwise, by the Pigeonhole Principle, there exists $k_0 < k_1$ such that $g(k_0) = g(k_1)$, in which case

$$\sum_{i=f(k_0)+1}^{f(k_1)} x_i = \sum_{j=k_0+1}^{k_1} y_j.$$

A-5 (3, 3, 0, 0, 0, 0, 0, 1, 16, 37, 147)

Solution 1. Observe first that the roots of $x^3 - 3x + 1$ can be isolated away from the given intervals; that is, there are sign changes in the intervals $[-2, -1]$, $[1/3, 1/2]$, $[3/2, 2]$. Hence the integrand is defined and continuous throughout.

Set

$$\begin{aligned} f(t) = & \int_{-100}^t \left(\frac{x^2 - x}{x^3 - 3x + 1} \right)^2 dx + \int_{\frac{1}{101}}^{1/(1-t)} \left(\frac{x^2 - x}{x^3 - 3x + 1} \right)^2 dx \\ & + \int_{\frac{101}{100}}^{1-1/t} \left(\frac{x^2 - x}{x^3 - 3x + 1} \right)^2 dx \end{aligned}$$

for $-100 \leq t \leq -10$. We wish to compute $f(-10)$. By the Fundamental Theorem of Calculus,

$$f'(t) = Q(t) + Q\left(\frac{1}{1-t}\right) \frac{1}{(1-t)^2} + Q\left(1 - \frac{1}{t}\right) \frac{1}{t^2}$$

where $Q(x) = ((x^2 - x)/(x^3 - 3x + 1))^2$. We find that $Q(1/(1-t)) = Q(1 - 1/t) = Q(t)$, so that

$$f(-10) = \int_{-100}^{-10} \left(\frac{x^2 - x}{x^3 - 3x + 1} \right)^2 \left(1 + \frac{1}{x^2} + \frac{1}{(1-x)^2} \right) dx.$$

But, noting that

$$\frac{1}{Q(x)} = \left(x + 1 - \frac{1}{x} - \frac{1}{x-1} \right)^2,$$

we see that the last integral is of the form $\int du/(u^2)$. Hence, its value is

$$-\frac{x^2 - x}{x^3 - 3x + 1} \Big|_{-100}^{10},$$

which is rational.

Solution 2. By the substitutions $x = -1/(t-1)$ and $x = 1 - 1/t$, the integrals over $[1/101, 1/11]$ and $[101/100, 11/10]$ are respectively converted into integrals over $[-100, -10]$. In the course of this it is seen that the function

$$\left(\frac{x^2 - x}{x^3 - 3x + 1} \right)^2$$

is invariant under each of the substitutions $x \rightarrow 1 - 1/x$ and $x \rightarrow -1/(x - 1)$ (which, in fact, are inverses of one another). Hence the sum of the three given integrals is expressible as

$$\int_{-100}^{-10} \left(\frac{x^2 - x}{x^3 - 3x + 1} \right)^2 \left(1 + \frac{1}{x^2} + \frac{1}{(x - 1)^2} \right) dx.$$

The solution continues as in the first solution.

Comment. It can be shown, more generally, that if $f(x)$ is a rational polynomial of degree at most 4, then

$$\int_{-100}^{-10} \frac{f(x)}{(x^3 - 3x + 1)^2} dx + \int_{1/101}^{1/11} \frac{f(x)}{(x^3 - 3x + 1)^2} dx + \int_{101/100}^{11/10} \frac{f(x)}{(x^3 - 3x + 1)^2} dx$$

is rational.

A-6 (0, 1, 0, 0, 0, 0, 1, 3, 3, 11, 36, 152)

Solution. We show that the conclusion holds with $r = 2 + \sqrt{3}$.

Assuming the result, we first derive the value of r . Observe that, asymptotically, the proportion of 2's in the first n terms is $1/r$. Thus, assuming there are about m 2's in the first $n \approx rm$ terms, there should be about $(r - 1)m$ 3's. These numbers give the approximate number of 3's in the intervals following the first n 2's, namely $2m + 3(r - 1)m = (3r - 1)m$. Hence, the proportion of 2's in the first $rm + (3r - 1)m = (4r - 1)m$ terms is $rm/(4r - 1)m = r/(4r - 1)$. So we want r to satisfy $1/r = r/(4r - 1)$, or $r^2 - 4r + 1 = 0$. Since r must exceed 1, $r = 2 + \sqrt{3}$.

Note that substituting $m = 0$ into the formula yields the first 2 and $m = 1$ the next. Assume the j th 2 is in position $1 + \lfloor r(j - 1) \rfloor$ for $j \leq m$. Solving $1 + \lfloor r(j - 1) \rfloor \leq m < 1 + \lfloor rj \rfloor$ yields $j = \lfloor m/(2 + \sqrt{3}) \rfloor + 1$ 2's among the first m numbers of the sequence. Thus the $(m + 1)$ st 2 is in position

$$\begin{aligned} & (m + 1) + 2(\lfloor m/(2 + \sqrt{3}) \rfloor + 1) + 3(m - \lfloor m/(2 + \sqrt{3}) \rfloor - 1) \\ &= 4m - \lfloor m/(2 + \sqrt{3}) \rfloor \\ &= \lfloor 4m - (2 - \sqrt{3})m \rfloor = \lfloor (2 + \sqrt{3})m \rfloor. \end{aligned}$$

The claim follows by induction.

B-1 (83, 29, 9, 0, 0, 0, 0, 0, 6, 10, 38, 32)

Solution. First, it is easily verified that

$$\frac{m}{1993} < \frac{2m + 1}{1993 + 1994} < \frac{m + 1}{1994},$$

so $n = 1993 + 1994 = 3987$ suffices. Now consider $m = 1992$ and suppose

$$\frac{1992}{1993} < \frac{k}{n} < \frac{1993}{1994}.$$

Since $x/(x+1)$ is strictly increasing for $x > 0$, we must have $k \leq n-2$ (note: $n > 1994$). However,

$$\frac{1992}{1993} < \frac{n-2}{n}$$

implies $3986 < n$, so $n \geq 3987$, completing the proof.

B-2 (89, 0, 1, 0, 0, 0, 0, 7, 3, 42, 65)

Solution. The probability that A wins is 0.

Clearly, A cannot win on the first turn. Assume B is to play, and that the total of announced numbers is T , and that A has cards x_1, x_2, \dots, x_k , and B has cards y_1, y_2, \dots, y_{k+1} . Because the integers $T + y_1, \dots, T + y_{k+1}$ have distinct remainders upon division by $2n+1$, at least one has a remainder other than $2n+1 - x_1, \dots, 2n+1 - x_k$. If B discards that y_i , it is impossible for A 's next discard to make the total divisible by $2n+1$. Therefore, A cannot win under optimal play by B .

B-3 (111, 16, 13, 0, 0, 0, 0, 4, 20, 10, 33)

Solution. The limit is $(5 - \pi)/4$ (that is, when $r = 5/4$, $s = -1/4$).

Note that the probability that x/y is exactly half an odd integer is 0, so we may safely ignore this possibility.

For any choice of x , the closest integer to x/y is even if either $x/y < .5$ or $2n - .5 < x/y < 2n + .5$ for some positive integer n .

The event $x/y < .5$, or $2x < y$, can occur only if $x < .5$. Thus its probability is

$$\int_0^{.5} (1 - 2x) dx = \frac{1}{4}.$$

For a positive integer n , the probability that $2n - .5 < x/y < 2n + .5$, i.e., that $2x/(4n+1) < y < 2x/(4n-1)$, is

$$\int_0^1 \left(\frac{2x}{4n-1} - \frac{2x}{4n+1} \right) dx = \frac{1}{4n-1} - \frac{1}{4n+1}.$$

Summing from $n = 1$ to ∞ , we get

$$\sum_{k=1}^{\infty} (-1)^{k+1} \frac{1}{2k+1} = 1 - \arctan 1 = 1 - \frac{\pi}{4}.$$

The total probability then is $1/4 + 1 - \pi/4 = (5 - \pi)/4$.

B-4 (0, 0, 0, 0, 0, 0, 0, 1, 1, 61, 144)

Solution. For $0 \leq x \leq 1$,

$$\begin{aligned} f(x) &= \int_0^1 g(t) K(x, t) dt \\ &= \int_0^1 \int_0^1 f(y) K(t, y) K(x, t) dy dt \\ &= \int_0^1 f(y) L(x, y) dy \end{aligned}$$

where

$$L(x, y) = \int_0^1 K(x, t)K(t, y) dt$$

for $0 \leq x \leq 1, 0 \leq y \leq 1$.

Similarly,

$$g(x) = \int_0^1 g(y)L(x, y)dy.$$

Since

$$\int_0^1 \frac{L(x, y)f(y)}{f(x)} dy = 1$$

and

$$\int_0^1 \left(\frac{L(x, y)f(y)}{f(x)} \right) \frac{g(y)}{f(y)} dy = \frac{g(x)}{f(x)},$$

for $0 \leq x \leq 1$, it follows that $g/f = c$, a constant. Then

$$\begin{aligned} g(x) &= cf(x) = c \int_0^1 g(y)K(x, y)dy \\ &= c^2 \int_0^1 f(y)K(x, y)dy \\ &= c^2 g(x). \end{aligned}$$

Therefore, $c = 1$.

B-5 (6, 1, 0, 0, 0, 0, 0, 1, 7, 65, 127)

Solution. Suppose there were 4 such points. Locate one point at the origin and let $\vec{v}_1, \vec{v}_2, \vec{v}_3$ be vectors from the origin to the other three. Since $\vec{v}_i \cdot \vec{v}_i$, and $|\vec{v}_i - \vec{v}_j|^2 = \vec{v}_i \cdot \vec{v}_i - 2\vec{v}_i \cdot \vec{v}_j + \vec{v}_j \cdot \vec{v}_j$, for $i \neq j$, are squares of odd integers we know $\vec{v}_i \cdot \vec{v}_i$ as well as $2\vec{v}_i \cdot \vec{v}_j$, for $j \neq i$ are integers congruent to 1 (mod 8).

Clearly no three points can be collinear. Hence $\vec{v}_3 = x\vec{v}_1 + y\vec{v}_2$ for some scalars x and y . Then

$$\begin{aligned} 2\vec{v}_1 \cdot \vec{v}_3 &= 2x\vec{v}_1 \cdot \vec{v}_1 + 2y\vec{v}_1 \cdot \vec{v}_2 \\ 2\vec{v}_2 \cdot \vec{v}_3 &= 2x\vec{v}_2 \cdot \vec{v}_1 + 2y\vec{v}_2 \cdot \vec{v}_2 \\ 2\vec{v}_3 \cdot \vec{v}_3 &= 2x\vec{v}_3 \cdot \vec{v}_1 + 2y\vec{v}_3 \cdot \vec{v}_2 \end{aligned} \tag{1}$$

Since \vec{v}_1 is not a scalar multiple of \vec{v}_2 ,

$$\det \begin{pmatrix} \vec{v}_1 \cdot \vec{v}_1 & \vec{v}_1 \cdot \vec{v}_2 \\ \vec{v}_2 \cdot \vec{v}_1 & \vec{v}_2 \cdot \vec{v}_2 \end{pmatrix} > 0$$

so that the first two equations in (1) have a unique rational solution for x, y , say $x = X/D$, $y = Y/D$, where X, Y , and D are integers. We may assume $\gcd(X, Y, D) = 1$. Then multiplying (1) through by D we have

$$\begin{aligned} D &\equiv 2X + Y \pmod{8} \\ D &\equiv X + 2Y \pmod{8} \\ 2D &\equiv X + Y \pmod{8} \end{aligned}$$

Adding the first two congruences and subtracting the third gives $2X + 2Y \equiv 0 \pmod{8}$, so that, by the third congruence D is even. But then the first two congruences force Y and X , respectively, to be even, a contradiction.

B-6 (2, 0, 0, 0, 0, 0, 0, 1, 2, 59, 143)

Solution. Say the numbers are a, b, c . First, we reduce to the case that exactly one of a, b, c is odd. Namely: (i) if two are odd, apply the rule with those two, and none is odd; (ii) if none is odd, divide all numbers by 2 and apply induction; (iii) if three are odd, apply the rule once, and exactly one is odd. Once exactly one is odd, this will remain so.

Say a is odd and b and c even. We aim to make the power of 2 dividing $b + c$ as large as possible. If b and c have the same number of factors of 2, then applying the rule to those two will yield both divisible by a higher power of 2, or one will have fewer factors of 2 than the other. Since $b + c$ is constant here, after a finite number of applications of the rule, b and c will not have the same number of factors of 2. Also, it is easy to see that, possibly after some additional moves, one has either $bc = 0$ (in which case one stops), or, the one of b and c divisible by the smaller power of 2 is also smaller; say it is b , so that $b < c$.

Case 1: $a > b$. Now work with a, b . Then a remains odd, b is doubled, and $b + c$ is divisible by a higher power of 2.

Case 2: $a < b$. Apply the rule first to a and b , and then to $b - a$ and c . (Note that $c > b > b - a$.) One obtains

$$\begin{array}{ccc} a & b & c \\ 2a & b - a & c \\ 2a & 2b - 2a & c - b + a \end{array}$$

Now the odd number is $c - b + a$, and the sum of the even numbers is $2b$, which has more factors of 2 than $b + c$.

Klosinski:
Department of Mathematics
Santa Clara University
Santa Clara, CA 95053

Alexanderson:
Department of Mathematics
Santa Clara University
Santa Clara, CA 95053

Larson:
Department of Mathematics
St. Olaf College
Northfield, MN 55057

Literacy in the Language of Mathematics

James O. Bullock

The American educational system has treated mathematics as skill in numerical manipulations, and has used the term “quantitative reasoning” to describe the application of mathematics to other areas of study. This serious misconception has severely hampered the ability of our students to comprehend important developments in scientific and philosophic thought. Mathematics can be more properly regarded as a form of language, developed by humankind in order to converse about the abstract concepts of numbers and space. In addition to the intellectual appeal of its intricate linguistic structure, the language should be appreciated for the richness of the literature composed with it.

The immediate cause for the recent concern about mathematics education is the widespread difficulty students apparently experience with this subject. While some educators have expressed an interest in improving pedagogy, students have tended to opt for a more practical approach whenever it has been open to them: avoiding the subject altogether. Predictably, the reaction of the educational system has been an attempt to force students to learn mathematics anyway. The mandate to hold students’ feet to the fire has been carried out with varying degrees of resolve, the debate over how much pain to inflict being dominated by two conflicting lines of reasoning. According to the first, there must be some minimum standards which can be applied to all students. In addition, since it builds character, everyone deserves their fair share of suffering. Those determined to adopt a more merciful attitude, on the other hand, often argue that the dose of unpleasantness should be limited to what is absolutely necessary. By this logic, mathematics courses need be required only when they are listed as prerequisites for courses that are required for some other reason.

In the course of such arguments, the question of mathematics itself has been relegated to a position of secondary importance. Implicit in this view is the premise that mathematics is a highly esoteric subject which, despite having some practical applications, possesses only tenuous connections to other areas of scholarship. Any consideration of the intellectual and aesthetic appeal possessed by the discipline has been dismissed as irrelevant for all, save the mathematicians themselves. Within most curricula, mathematics has been so thoroughly dissociated from all other subjects that students generally encounter little evidence that these assumptions might be mistaken. Physical science, for example, is nearly always taught as if students know little or nothing of mathematics. Similarly, students are expected to learn biology without any reliance upon the concepts and formalism of the physical sciences. Many statisticians now insist that their subject is something quite apart from mathematics, so that statistics courses do not require any preparation in mathematics. Some computer scientists have gone so far as to suggest that if you understand the mathematics of their machines, it is not necessary to understand the mathematics of natural phenomena. To the students, nothing seems to be

amiss. They observe that the factors which determine whether a course is classified as hard or soft, general or specialized, are the complexity of the problem assignments and the thickness of the textbook. As the values of these parameters increase, the demand for *proficiency* with specific mathematical techniques also increases. Across the entire spectrum of course offerings, however, students are exposed only to the vaguest notions of what we mean by *rigor*.

A careful consideration of the nature of mathematics as an intellectual activity reveals the utter incongruity of this situation. Not only have our students failed to appreciate the beauty of mathematics, they have little grasp of the profound insights about the natural world which mathematics has made possible. Rather than a mere intellectual curiosity or useful skill, mathematics is an important facet of the most distinctive capability of the human species.

MATHEMATICS IS A LANGUAGE. As a general definition, mathematics could be called the study of numbers and space. In order to talk about such things, mathematicians first had to devise an appropriate vocabulary and alphabet. It is apparent that the objects defined by mathematicians are entirely *abstract*, and can never actually be observed in any way except by the human imagination. Although many students seem to be disturbed by this fact, mathematics is comprehensible precisely *because* it is abstract. We are not dependent upon observation to know that lines are completely straight and parallel lines never meet. It is possible to make these assertions with confidence because mathematics was not discovered, it was invented. It must also be remembered that imagination and abstraction are universal parts of the human experience, and not the exclusive domain of the mathematician. The idea of a line or a point is no more abstract than the idea of loyalty or freedom. Mathematics differs from other languages not because of its inclusion of abstraction, but because of its complete detachment from the complications of what we experience by direct observation.

If mathematics consisted only of new words and symbols, it could properly be considered as an extension of existing language. The reason mathematics is a new and separate language is that it also has its own syntax and grammar. Having devised both vocabulary and rules for the language, mathematicians seek to discover what things are possible or not possible to say about numbers and space. In this ideal, ordered world of the imagination, it is possible to apply relentless logic to any questions that arise. Euclid's geometry should be appreciated for the beauty which can result from this process. His system of axioms, definitions, theorems, and constructions clearly lays out both the vocabulary and the rules for conversing about his chosen topic. By limiting his considerations to fanciful but specific abstractions of his imagination, he was able to produce an indisputably logical and thorough exposition. If one understands the language, one will not only know what a triangle is, but exactly what can be said about triangles. The lexicon and grammar of language are quite precise; the conclusions are quite inescapable.

FROM THE LINGUISTIC TO THE LITERARY. The natural sciences seek to explain the behavior of the things we encounter in nature. Experimental observation is the court of highest appeal for all such explanations. What then is the role of the abstract, detached language of mathematics in experimental science? The answer is not different than that for the roles of language in other forms of scholarship. The scientist uses the language of mathematics to construct *metaphors* which represent insights into the workings of nature. The use of metaphors is as commonplace in science as it is in poetry. We speak of radio "waves", subatomic

“particles”, and celestial “bodies”. What distinguishes the mathematical metaphor is the extraordinary power of this language to uncover implications of the underlying idea. Calling the earth a sphere is an extremely powerful statement, for we know a great deal about the things we can say about spheres. The fact that most scientists prefer to call such constructions “models” or “laws” does not change their essential character. It is just as metaphorical to call the world a sphere as it is to call it a stage.

GETTING THE PICTURE. It is generally agreed that the object of studying science is to understand its fundamental concepts, but how does one come to understand a mathematical metaphor? It is widely believed that the first step is to find a way to remove the mathematical details so that the essential picture can be seen unobscured. Mathematics, then, need be introduced only if one is confronted with a specific problem that requires a numerical answer. Many educators are fond of using the term “quantitative reasoning” as a substitute for “mathematics”. The idea that mathematical questions can be answered qualitatively without actually using any mathematics represents one of the most serious misconceptions about the nature of this discipline. Mathematics is not a way of hanging numbers on things so that quantitative answers to ordinary questions can be obtained. It is a language that allows one to think about extraordinary questions. Saying that the earth has a round shape means only that it has no edges. This non-mathematical picture is not simply “qualitative”, “verbal”, or “intuitive”; it is primitive and empty. If we wish to construct a meaningful metaphor about the shape of the earth, we must use the language of shapes, which is mathematics. To those prepared to examine its implications, even the mathematically simple picture of the spherical shape has a great richness when used as a metaphorical device. Getting the picture does not mean writing the formula or crunching the numbers, it means grasping the metaphor. Without mathematics, one cannot even read the words.

READING CRITICALLY. Students of the physical sciences are often left with the impression that facility in solving mathematical problems is all that is necessary to understand scientific ideas. After all, the entire subject of classical physics can be reduced to seven compact, simple-looking equations, and most study time in this subject is spent working problems. This view, however, has the same basic flaw as the belief that mathematics can be replaced by less abstract language. Just as excluding numerical calculations from an argument does not eliminate the need for mathematics, one cannot bring mathematics to bear on a question by simply assigning numbers and symbols to its elements. Metaphors do not have answers, they have implications. When we struggle to find solutions to mathematical problems, we are exploring the specific implications of a writer’s metaphors in particular physical circumstances.

In physics, as in all other intellectual disciplines, understanding derives from critical thought, not just hard work. The equations of electrodynamics, for example, are important not simply because they give rise to interesting and complicated mathematical problems. Maxwell’s picture of flowing force fields which diverge and swirl according to precise mathematical relationships is an incredibly rich metaphor. In order to understand this literature, it must not only be read, it must be pondered carefully. By thoughtful consideration of wisely chosen examples, you can begin to discover the things about different physical situations that have important influences on the outcomes one observes. You begin to understand the

metaphor. As Paul Dirac put it: "I understand what an equation means if I have a way of figuring out the characteristics of its solution without actually solving it." This is what we mean by physical insight; it is an intuition based upon critical analysis of key questions. Einstein's surmise that magnetism is fundamentally a relativistic effect of electricity, one of the truly remarkable insights of human history, came directly from his contemplation of Maxwell's equations. Building such intuition may require considerable effort. Electricity and magnetism are more difficult to understand than the shape of the earth, and Maxwell's metaphor is complex in language and construction. On the other hand, one could spend a lifetime working out solutions to very complex and difficult problems in electrodynamics and never discover the possibility of electromagnetic radiation, even though Maxwell's equations directly imply its existence. A mason may indeed build great strength in his arm muscles, but his goal is to build the house. Likewise, the goal of the scholar is not technical dexterity, but insight.

ANALYZING MATHEMATICAL LITERATURE. While it is important to choose interesting examples to think about, as a practical matter we are still left with mathematical problems to solve. In asserting that mathematics is an idealized abstraction which lends itself to elegant, logical treatments, we have overstated our case in several particulars. First, not all mathematical problems have solutions. In fact, the only way a mathematical system can be made to give uniformly consistent answers is for some questions to be left completely unanswered. Even if solutions are known to exist, it is sometimes true that there can be no general method for finding them. (The most famous example is the problem of trisecting an angle.) Such philosophical difficulties with the positivist viewpoint pale in comparison to the practical difficulties in finding solutions, even when you know that they exist and that you should be able to find them. Specific mathematical problems, no matter how well posed, often lead one into an impenetrable thicket from which it can be exceedingly difficult to extract solutions by any means. In attempting to understand science, one should not let these kinds of mathematical difficulties take over. What the scientist wants to know is whether the original metaphor can reveal anything about what will be expected to happen in a particular situation. Often, a good approximate solution can lead out of the quagmire toward great insight. The concepts of field lines, resistance, and capacitance are examples of insightful approximations of Maxwell's equations which allow one to picture and analyze the behavior of certain physical systems by avoiding purely mathematical entanglements. Hence, the use of these heuristic devices is another way to achieve the kind of understanding to which Dirac referred. One must never lose sight of the goal of this process. By itself, an individual solution or the particular technique used for finding it is important only insofar as it provides clues about the nature of the metaphor. It is in the original equations, not in a catalog of solutions, that the complete law, the entire metaphor, is to be found. The goal is to grasp the metaphor, not just to compile a catalog.

THE NATURE OF KNOWLEDGE. Finally, we have to consider that the metaphors themselves may have limited usefulness. Newton's laws of motion have turned out to need corrections in order to be consistent with the theory of relativity. The problems with Maxwell's laws are more serious: they are internally inconsistent at some points. Even the most imaginative metaphor cannot bring every facet of truth into plain view. In non-mathematical literature, we are used to the fact that meanings become distorted when ideas are stretched too far. (For example, the

cartoonist Johnny Hart observed some years ago that the trouble with a melting pot is that the bottom gets burnt and the scum comes to the top.) In interpreting mathematical language in the scientific literature, we tend to become confused by the Platonistic view that mathematical objects themselves actually exist. Regardless of whether this assertion about mathematics is philosophically correct, the use of this language does not imbue one's ideas about nature with an independent existence. The use of mathematical objects is still a metaphorical device. Mathematical metaphors are not different from other forms of human understanding. They may provide some significant insights, but are not a holy grail of ultimate truth. This should come as no surprise. We know that the earth is not *exactly* a sphere. Furthermore, the holy grail is a heretical idea, not just to Christian theology, but to Western philosophy as well. Despite our desires to the contrary, nature is not obliged to limit its complexity to the confines of the human imagination. Despair is not an appropriate reaction to this revelation. The importance of the insights provided by any single idea is not diminished because it fails to explain everything else in nature. In addition, our capabilities are not forever fixed. Mathematicians have continued to explore new territories in the abstract landscape they have created, and certainly do not believe that this is a task which will one day be completed. If all of human understanding is based on metaphors, then extending the reach of the languages we use to construct them can greatly expand the boundaries of our imagination. We have cited but a scant few of the important advances in the natural sciences which have been made possible by the development of mathematics. In a fuller accounting, we would need to discuss its influences in philosophy and in the social sciences, as well as the fresh insights provided by new mathematics into old conundrums. Mathematics, therefore, does not simply represent a narrow but important class of intellectual achievement. Its distinct and prominent place in the realm of human language and literature make it an essential foundational element of education.

HOW DOES ONE BECOME LITERATE? As educators, we seek to engage students in the exploration of literature, in the understanding of other people's ideas, and in the expression of their own views. If the starting place of this process is *language*, its goal must be *literacy*. This goal can be attained only through thoughtful reading and critical analysis. Having seen that all forms of language share the same basic literary device, we can draw rather extensive parallels between the meaning of literacy in the language of mathematics and that which applies to other languages. The following list of assertions about the nature of literacy hardly requires explanation or justification. When applied to mathematics education, however, they imply that our current approaches are grossly ill conceived, and require wholesale revision.

Individuals can learn more than one language. The assertion that mathematics requires a very special innate talent is often used as an excuse for not learning this subject and a reason for not trying to teach it. This sort of condescension must not be permitted in the educational community. Mathematics is not so different from other expressions of the human intellectual capacities for communication and imagination. The fact one can learn to say things in mathematical language that cannot be said in other languages is beside the point, as is the possibility that different regions of the brain may be used. Mathematics is not the exclusive province of either the gifted or the deranged, it is for all who would seek to be truly educated.

One must learn to read before trying to study literature. This point seems so obvious that it is surprising that it is violated so systematically, especially in engineering and science curricula. Unsuspecting students are routinely confronted by passages of unintelligible mathematics. The accompanying explanation, if one is to be found, is typically cursory. Usually, it is unclear whether the mathematical interlude is given as an aside, whether it is a point of central importance, or whether it is something the students were supposed to have known already. Often, such narratives become more obscure as they progress. In reading any form of literature, one may occasionally find it necessary to consult a dictionary, and footnotes can be very helpful in clarifying points of grammar or usage. When one has to look up every other word and the footnotes begin to swallow up the text, however, a reader has very little chance of putting together the meaning of even a single complete sentence. If students are given the chance to study enough mathematics *first*, it can be startling how facily they are able to grasp the central ideas of other disciplines.

Vocabulary is not enough. We all remember vocabulary tests from grade school: look up the word in the dictionary, memorize its definition, learn how to spell it, and go on to the next word. We compile the same sorts of lists for mathematics students: quadratic equations, logarithms, antiderivatives, Bessel functions, etc. Furthermore, we forbid our students to use a dictionary to look up the unfamiliar mathematics they may encounter in their studies. We insist that they commit the entire dictionary to memory. The difficulty with this approach is that knowing the formal definitions of a large number of words is no guarantee that one can actually say anything comprehensible. One must obtain practice in expressing complete thoughts with language. In mathematics, this means constructing examples which deal with specific objects and events; that is, applications. It does not matter so much whether the objects are observable or abstract, just as it does not matter whether the words are in the form of prose or verse. What matters is that words must be given a context if we are to be enriched by understanding their meanings.

Conversational fluency is of limited value. We spend a great deal of time telling our students “this is how you solve this kind of problem.” To divide by a fraction, for example, you just invert and multiply. This is just like saying “when you answer the telephone, pick up the receiver and say ‘hello’.” This might seem to be a step forward from the vocabulary lists, but most often it actually represents a retreat. You do not even have to know what the individual words mean in order to be able to fire back the correct response. Once you have learned the drill, you will never have to think about it again. In short, this is *training* rather than *education*. Whether or not any sort of training has a proper role in university curricula is a matter of longstanding controversy, but in this case, the point is moot. We have trained our students to become highly proficient in answering questions they will never be asked again. Outside the classroom, it is extremely unlikely that you will be asked to “solve for x ”. The harm we inflict by insisting on this sort of raw skill development goes far beyond simply wasting our students’ time, however. Since they never really understood the meaning of the answers they were trained to give, they will be unable to answer the questions that do come up. Because they come to perceive mathematics as a collection of specialized skills, rather than a way of thinking and talking, our students are lead to conclude that mathematics is either worthless or impossible to master. Imagine what we would think if they came to this conclusion about, say, Spanish. There was a line in one of our high school

dialogues that went: “they-always-serve-meatballs-in-the-cafeteria-on-Wednesday.” Since I have never had the opportunity to use Spanish to tell someone this, am I to conclude that Spanish is worthless for communicating with my fellow humans, or that it is too much to expect that I will ever be able to learn how to say anything really useful? By offering training rather than education in mathematics, we have produced a legion of mathematical sophomores, possessed of an extensive but superficial knowledge. Breaking this cycle is one of the most important pedagogical challenges in mathematics education.

Writing is not the same as penmanship. Whereas penmanship used to be considered a highly valuable skill, most of us would now refuse to accept a handwritten document, no matter how legible or graceful the hand. We would not say that a manuscript had been “written” by a machine just because it was typed. In the case of mathematics, our attitude is exactly the opposite. We insist on work which has been done by hand, and disallow all forms of mechanical or electronic assistance because it amounts to cheating. Whether the students are in college learning calculus or in the third grade learning multiplication, we insist that they concentrate on developing their skill at the mechanical manipulation of symbols required to arrive at the desired product. The time for us to abandon this antiquated and misguided attitude is long past. We are wasting our students’ time by teaching them how to write with a quill pen when they should be using a typewriter. The use of modern technology is beneficial to mathematical undertakings and literary tasks alike. One’s attention can be focused on the conceptual content of what is being written, rather than the manipulative processes required to produce the writing itself. The skills of using the typewriter, the calculator, and the computer are useful ones. The value of being able to add large columns of number in one’s head or remember the Fourier transforms of complicated functions is dubious at best. More to the point, in the outside world, it will be considered inappropriate to waste time performing such tasks by hand.

Complicated is not the same thing as sophisticated. Nearly everyone has had experience with long, tedious problems involving only elementary arithmetic. There is no trick to constructing similar problems at any level of mathematics. When a problem appears to be difficult, it is important to distinguish whether obtaining the solution requires new insights, a bit of cleverness, or just plain drudgery. The simple problems are not the only ones that are worth solving, but it is appropriate to offer some justification for tackling difficult ones.

The study of grammar can be overdone. While systematic consideration of mathematical grammar is generally ignored in lower level courses in mathematics, it often overwhelms everything else in advanced courses. For someone who is studying mathematics in order to do a bit of advanced reading, it is not always necessary to grasp all of the subtleties of mathematical arguments in order to appreciate the usefulness of their conclusions. Some such problems, like proving the existence of the derivative, turn out to be quite formidable. It is sometimes enough to understand the kinds of things mathematicians worry about, but to leave the actual worrying to them.

It is possible to be fluently ignorant. The ability to read the English language does not imply any knowledge of Shakespeare’s plays, nor is an understanding of the playwright’s ideas derived simply from hearing his words. Further, because a piece of writing is grammatically correct and draws from a large vocabulary is no

indication that it contains any worthwhile ideas. Similarly, the ability to perform mathematical gymnastics is not necessarily indicative of any underlying understanding of either the literature of mathematics or the nature of the language. Mathematical exercises should always be chosen in such a way as to provide the student with a keener insight into great concepts, not just practice in manipulations. Education is not just learning to read and write. The questions of what is read and what is written about are central to any practical definition of literacy.

Literature must be studied from original sources. The application of this principle to science and mathematics is often misunderstood. We are not suggesting that the only proper way to learn Newtonian physics is to study the *Principia*. Ideas and arguments presented in the language of mathematics can be reproduced very concisely. When that part of the exposition written in other languages is paraphrased or even condensed, little or no loss of meaning need occur as long as the mathematics is preserved. On the other hand, major distortions of perception will inevitably result if mathematics is replaced by non-mathematical pictures or lists of specific examples. It is essential to present the original formulation, the equations, which define the author's metaphor. This is true even if the students' command of the language is limited, and they are able to grasp only some of the simpler implications of the metaphor. Having seen the original idea in its entirety, they will understand the underlying basis of however much they are ready to learn. Furthermore, they can approach the subject again in the future without having anything to unlearn.

The accomplishments of the past cannot be dismissed. We have become accustomed to considering pre-modern science as so much primitive nonsense: domed sky, flat earth, flies erupting spontaneously from dead meat, et cetera. In the humanities, such an attitude would appear silly. The works of Shakespeare did not relegate those of Sophocles to the dust bin. Because mathematics is generally, and quite incorrectly, categorized with science rather than with philosophy, most people, including our students, tend to think that the mathematics being taught in the present day represents modern developments. In fact, most college students are exposed to very few mathematical ideas which originated since the fifteenth century. The accomplishments of the early Greek, Hindu, and Islamic scholars are of more than historical interest. They were not made obsolete by subsequent developments, but rather formed the basis for those very advances. What is distressing about the present state of education is that so few students have even an inkling of the progress made in mathematics during the last half millennium.

There is no substitute for learning how to read. The reason mathematics was devised in the first place was the inability of existing language to deal with this subject matter. Newton, for example, found it necessary to invent the calculus in order to develop and express his ideas. Trying to understand Newton without calculus is *not* like trying to understand Sophocles without Greek, it is like trying to understand Sophocles without *words*. One could represent *Antigone* as a series of pictures. If done with sufficient skill, the major points of the story line could be made evident; but the play is not just a story about some people who end up dead under tragic circumstances. To pretend that one can understand Newton without using his mathematics would be equally delusional. At best, one might hope to be trained in solving certain types of problems involving throwing things up in the air or sliding them down inclined planes. With few exceptions, such training will prove

useless in the long run. Visual images, no matter how important they may be to human communication, do not constitute a form of language. Creating a picture book is not an act of translation; and, despite the aphorism to the contrary, no number of pictures can match the power of the written or spoken word. Mathematics, on the other hand, is precisely a form of language. If we truly wish to understand the things that have been said with this language, we must eventually take up some books with words in them.

Illiteracy has unfortunate consequences. In demanding picture books, one chooses illiteracy and is cut off from some of the most important ideas in the history of western civilization. This is a decision that should not be taken lightly by anyone who truly desires an education. That Americans seem unaware of the consequences of such a decision is hardly surprising, since they do not make it for themselves. If it makes us feel better about our own ignorance to say that we know that the earth is spherical, it is only because we were never told that the educated people of the world had known that for at least 1800 years prior to the time Columbus set sail. Americans may not think that the earth is flat, but most do believe that Michael Jordan can hang in the air and change direction in mid-flight. Only the ignorant could be convinced of this. The educated have known better for three hundred years. Our own educational system has given out nothing but picture books, and as a result, our society has been left illiterate and vulnerable to manipulation. We have told the public that they should leave the technical stuff to the experts, that they are not smart enough to understand anyway. They have been convinced that they are stupid. They are not stupid, they have been cheated.

Literacy is the goal of education. The study of mathematics not only allows one to read and understand the work of others, it also increases one's own powers of thought, imagination, and expression. That so many regard this subject as frightening, boring, or otherwise unpleasant represents a disappointing failure of our educational efforts. Literacy in mathematics is not simply a question of how much or for whom. To improve our present condition will require substantive reform across the spectrum of curricula. It will not be enough for our departments of mathematics to teach our students how to read. All the rest of us must see to it that they do read, that what they read is worthwhile and important, and that they are able to comprehend what they read. Becoming literate requires a *real* education, the kind that does not become out of date, but prepares one for a lifetime of learning. Just as there is no aspect of the human experience unworthy of serious study, there is no student unworthy of a genuine education. For once, a popular buzz word has captured the true essence of a national problem. Our difficulty is not that education has failed to keep pace with rapidly changing knowledge, it is that we have misplaced the goal of all true education: literacy.

REFERENCES

1. Adler, M. J. and C. Van Doren. 1972. *How to Read a Book*. Simon and Schuster. New York.
2. Feynman, R. P., R. B. Leighton, and M. Sands. 1964. *The Feynman Lectures on Physics*. Addison-Wesley Publishing Co. Reading, Massachusetts.
3. Whorf, B. L. 1956. *Language, Thought, and Reality*. Technology Press and John Wiley & Sons, Inc. New York.

*Department of Physiology
University of Missouri–Columbia
Columbia, MO 65212*

Fractional and Trigonometric Expressions for Matrices

Goro Shimura

1. INTRODUCTION. We start with two elementary facts:

(I) *Every rational number x can be expressed as a quotient c/d with two integers c and d which have no common divisor greater than 1. Moreover, if $x = c'/d'$ is another such expression, then $(c', d') = \pm(c, d)$.*

(II) *Every real number x can be expressed as a quotient c/d with two real numbers c and d such that $c^2 + d^2 = 1$. Moreover, if $x = c'/d'$ is another such expression, then $(c', d') = \pm(c, d)$.*

One can add some geometric flavor to the latter statement by using $\sin \theta$ and $\cos \theta$ instead of c and d .

There is a certain parallelism between these two statements, but I guess most readers will wonder if it is really essential or merely rhetorical. This point aside, we can pose the question of whether the same types of statements can be made for matrices, which is the meaning of the title of this article. Considering only the first type for the moment, we can ask:

(A) *Given a matrix x with rational entries, can one find matrices c and d with integral entries so that $x = d^{-1}c$ and that this is an expression “reduced to the lowest terms”? If the answer is affirmative, to what extent is the pair (c, d) unique?*

Here we don't have to assume that x is square; if x is an $m \times n$ -matrix, then c must be of the same shape and d invertible and $m \times m$. Similarly we can ask for an expression $x = cd^{-1}$ with c and d of appropriate shapes. There is of course a trivial answer: take d to be the scalar which is the least common denominator of the entries of x . But look at the equality

$$\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 7 & 1 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix}^{-1} \begin{pmatrix} 21 & 3 \\ 0 & 10 \end{pmatrix}.$$

Which side is more appropriately called the quotient reduced to “the lowest terms”? Clearly the left-hand side looks more “reduced.”

This example will tell that perhaps the answer to (A) should be formulated in terms of *elementary divisors* of integral matrices, and in fact (A) can be answered easily that way, as will be shown below. However, our aim is not only to give an answer to (A), which is quite elementary, but also to supply a certain conceptual background for this type of question. At the same time, we shall show that (II), as well as its matrix version, can be understood in the same framework. In the last two sections we shall give higher-dimensional analogues of $\cos \theta$ and $\sin \theta$, as well as their hyperbolic counterparts, in the context of statement (II), which are closely related to the decomposition of the type $G = K \cdot \exp(\mathfrak{p})$ for *compact and noncompact* Lie groups G . Our theorems or lemmas in the other sections can hardly be

called new, but it seems that some of them have never been stated in the forms as we present them in this article.

2. FRACTIONAL EXPRESSIONS FOR RATIONAL AND REAL MATRICES. Let us first make some notational conventions. For an associative ring R with identity element, $M_n^m(R)$ denotes the module of all $m \times n$ -matrices with entries in R and $GL_n(R)$ the group of all invertible elements of the ring $M_n(R)$. The zero element of $M_n(R)$ is denoted by 0_n^m or simply by 0 , the identity element of $M_n(R)$ by 1_n or simply by 1 , and the transpose of a matrix X by tX . For square matrices A_1, \dots, A_r we denote by $\text{diag}[A_1, \dots, A_r]$ the square matrix with the A_i in the diagonal blocks and 0 everywhere else. As usual, \mathbf{Z} , \mathbf{Q} , and \mathbf{R} denote the ring of integers, the field of rational numbers, and the field of real numbers.

We now recall

Lemma 1. *Given $X \in M_n^m(\mathbf{Q})$ of rank r , there exist $A \in GL_m(\mathbf{Z})$, $B \in GL_n(\mathbf{Z})$ and positive rational numbers e_1, \dots, e_r such that $e_{i+1}/e_i \in \mathbf{Z}$ for all $i < r$ and*

$$AXB = \begin{pmatrix} E & 0_{n-r}^r \\ 0_r^{m-r} & 0_{n-r}^{m-r} \end{pmatrix}, \quad E = \text{diag}[e_1, \dots, e_r]. \quad (2.1)$$

Moreover, the e_i are uniquely determined by X .

This is usually stated only for $X \in M_n^m(\mathbf{Z})$, but for $X \in M_n^m(\mathbf{Q})$ if we take $g \in \mathbf{Z}$, $g > 0$, so that $gX \in M_n^m(\mathbf{Z})$ and apply the theorem to gX , then we immediately obtain the result in the above form. The numbers e_1, \dots, e_r are called the *elementary divisors* of X .

We call an element X of $M_n^m(\mathbf{Z})$ *primitive* if $\text{rank}(X) = \text{Min}(m, n)$ and the elementary divisors of X are all equal to 1. If $m = n$, clearly X is primitive if and only if $X \in GL_n(\mathbf{Z})$.

Lemma 2. *Let $X \in M_n^m(\mathbf{Z})$ and $m \leq n$. Then the following conditions on X are mutually equivalent:*

- (1) X is primitive.
- (2) There exists an element Y of $M_n^{n-m}(\mathbf{Z})$ such that $\begin{pmatrix} X \\ Y \end{pmatrix} \in GL_n(\mathbf{Z})$.
- (3) There exists an element W of $M_m^n(\mathbf{Z})$ such that $XW = 1_m$.
- (4) If $C \in M_m^1(\mathbf{Q})$ and $CX \in M_n^1(\mathbf{Z})$, then $C \in M_m^1(\mathbf{Z})$.
- (5) For every $l \in \mathbf{Z}$, $l > 0$, if $C \in M_m^l(\mathbf{Q})$ and $CX \in M_n^l(\mathbf{Z})$, then $C \in M_m^l(\mathbf{Z})$.

Proof: Put $s = n - m$. If X is primitive, our definition implies that $AXB = \begin{pmatrix} 1_m & 0_s^m \end{pmatrix}$ with $A \in GL_m(\mathbf{Z})$ and $B \in GL_n(\mathbf{Z})$. Let $Y = \begin{pmatrix} 0_s^s & 1_s \end{pmatrix} B^{-1}$. Then

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} A^{-1} & 0 \\ 0 & 1_s \end{pmatrix} B^{-1} \in GL_n(\mathbf{Z}),$$

which proves that (1) \Rightarrow (2). With Y as in (2) put $\begin{pmatrix} X \\ Y \end{pmatrix}^{-1} = \begin{pmatrix} W & Z \end{pmatrix}$ with $W \in M_m^n(\mathbf{Z})$ and $Z \in M_s^n(\mathbf{Z})$. Then $XW = 1_m$, and hence (2) \Rightarrow (3). With W as in (3) let $C \in M_m^1(\mathbf{Q})$ and $CX \in M_n^1(\mathbf{Z})$. Then $C = CXW \in M_m^1(\mathbf{Z})$, and hence (3) \Rightarrow (4). Decomposing C of (5) into row vectors, we easily see that (4) \Leftrightarrow (5). Finally, assume (4). If $\text{rank}(X) < m$, then there is a vector $v \in M_m^1(\mathbf{Q})$, $v \neq 0$, such that $vX = 0$. Changing v for its suitable rational multiple if necessary, we may assume that $v \notin M_m^1(\mathbf{Z})$. This cannot happen under (4). Thus $\text{rank}(X) = m$. Now take A ,

B , E , and e_1, \dots, e_r as in Lemma 1. Then $E^{-1}AX = (1_m \ 0)B^{-1} \in M_n^m(\mathbf{Z})$. Then (5) implies that $E^{-1}A \in M_m^m(\mathbf{Z})$, and hence $E^{-1} \in M_m^m(\mathbf{Z})$. Therefore $e_i = 1$ for every i , which proves that (4) \Rightarrow (1) and our proof is complete.

Now an answer to (A) can be given as follows:

Theorem 1. *Given $x \in M_n^m(\mathbf{Q})$, the following assertions hold:*

(1) *There exist $c \in M_n^m(\mathbf{Z})$ and $d \in M_m^m(\mathbf{Z}) \cap GL_m(\mathbf{Q})$ such that $(c \ d)$ is primitive and $x = d^{-1}c$.*

(2) *If c and d are as in (1) and if $x = d'^{-1}c'$ with $c' \in M_n^m(\mathbf{Z})$ and $d' \in M_m^m(\mathbf{Z}) \cap GL_m(\mathbf{Q})$, then $c' = hd$ and $d' = hd$ with $h \in M_m^m(\mathbf{Z})$. In addition if $(c' \ d')$ is also primitive, then $h \in GL_m(\mathbf{Z})$.*

(3) *Let e_1, \dots, e_r be the elementary divisors of x and for each i let $e_i = f_i/g_i$ with relatively prime positive integers f_i and g_i . If c and d are as in (1), then the elementary divisors of c are f_1, \dots, f_r and the elementary divisors of d are $1, \dots, 1, g_r, \dots, g_1$ with 1 repeated $m - r$ times.*

Proof: Assuming the existence of c and d as in (1), suppose $x = d'^{-1}c'$ with $c' \in M_n^m(\mathbf{Z})$ and $d' \in M_m^m(\mathbf{Z}) \cap GL_m(\mathbf{Q})$; put $h = d'd^{-1}$. Then $d' = hd$, $c' = d'x = hc$, and so $h(c \ d) = (c' \ d') \in M_n^m(\mathbf{Z})$. By (5) of Lemma 2 we have $h \in M_m^m(\mathbf{Z})$. If $(c' \ d')$ is also primitive, then exchanging $(c \ d)$ for $(c' \ d')$, we find that $h^{-1} \in M_m^m(\mathbf{Z})$, and hence $h \in GL_m(\mathbf{Z})$. This proves (2). To prove (1) and (3), we apply Lemma 1 to x to find an expression $axb = \begin{pmatrix} e & 0 \\ 0 & 0 \end{pmatrix}$ with $a \in GL_m(\mathbf{Z})$, $b \in GL_n(\mathbf{Z})$, $e = \text{diag}[e_1, \dots, e_r]$, $0 < e_i \in \mathbf{Q}$, $e_{i+1}/e_i \in \mathbf{Z}$. Put $e_i = f_i/g_i$ as in (3) and $d = \text{diag}[g_1, \dots, g_r, 1_{m-r}]a$, $f = \text{diag}[f_1, \dots, f_r]$, $q = \begin{pmatrix} f & 0 \\ 0 & 0 \end{pmatrix}$, and $c = qb^{-1}$, where the zeros in q are arranged so that $q \in M_n^m(\mathbf{Z})$. Then $x = d^{-1}c$ and

$$(c \ d) \begin{pmatrix} b & 0 \\ 0 & a^{-1} \end{pmatrix} = \begin{pmatrix} f_1 & & 0 & 0 & g_1 & & 0 & 0 \\ & \ddots & & & & \ddots & & \\ 0 & & f_r & 0 & 0 & & g_r & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1_{m-r} \end{pmatrix}.$$

We easily see that the right-hand side is primitive (by checking (4) of Lemma 2, for example), and hence $(c \ d)$ is primitive since $\text{diag}[b, a^{-1}] \in GL_{m+n}(\mathbf{Z})$. This proves (1). Put $k_i = e_{i+1}/e_i$. Then $k_i f_i g_{i+1} = g_i f_{i+1}$. Since f_i and g_i are relatively prime, we see that $f_i | f_{i+1}$, and similarly $g_{i+1} | g_i$. Therefore we obtain (3).

Remarks. (R1) If $(c \ d)$ is as in Theorem 1, then (3) of Lemma 2 guarantees elements $a \in M_m^n(\mathbf{Z})$ and $b \in M_m^m(\mathbf{Z})$ such that $ca + db = 1_m$. Thus, as in number theory, $(c \ d)$ may be called *relatively prime*. It should be noted that as can be seen from the example $x = \text{diag}[2, 1]^{-1} \text{diag}[1, 2]$, the maximum elementary divisors of c and d may not be relatively prime.

(R2) The expression $x = cd^{-1}$ can be treated in the same fashion by making obvious modifications. For example, in this case we assume that $\begin{pmatrix} c \\ d \end{pmatrix}$ is primitive.

(R3) Clearly all the above results can be extended to the case of matrices with entries in the field of quotients of a principal ideal domain, or in a central division algebra over a p -adic field. This comment applies also to (1) of Theorems 3, 4, 5 and Lemma 6 below.

Let us now turn to real matrices. In fact, we shall treat not only real matrices but also complex and quaternion matrices, since all these can be handled easily by the same technique. To make our exposition uniform, we let \mathbf{K} denote any one of the following three objects: the real number field \mathbf{R} , the complex number field \mathbf{C} , and the division ring of Hamilton quaternions \mathbf{H} . For $X = (x_{ij}) \in M_n^m(\mathbf{K})$ we define $X^* \in M_m^n(\mathbf{K})$ by $X^* = (y_{ij})$ with $y_{ij} = \bar{x}_{ji}$, where \bar{x} is the image of x under complex conjugation or quaternion conjugation. We put also

$$U_n(\mathbf{K}) = \{\alpha \in GL_n(\mathbf{K}) \mid \alpha\alpha^* = 1_n\},$$

$$S_n(\mathbf{K}) = \{X \in M_n^n(\mathbf{K}) \mid X^* = X\},$$

and call an element X of $S_n(\mathbf{K})$ *positive definite* (resp. *nonnegative*) if $y^*Xy > 0$ (resp. $y^*Xy \geq 0$) for every $y \in M_1^n(\mathbf{K})$, $\neq 0$. If $\mathbf{K} = \mathbf{R}$, then $X^* = {}^tX$ and $U_n(\mathbf{K})$ is the orthogonal group of degree n . For $X, Y \in S_n(\mathbf{K})$ we write $X > Y$ and $Y < X$ if $X - Y$ is positive definite.

Lemma 3. *If $y \in M_n^m(\mathbf{K})$, then yy^* is nonnegative. Conversely, every positive definite element s of $S_n(\mathbf{K})$ can be written in the form $s = r^2$ with a positive definite element r of $S_n(\mathbf{K})$ and also in the form $s = aa^*$ with an upper triangular matrix a in $M_n^n(\mathbf{K})$ whose diagonal elements are positive real numbers. Both r and a are unique for s .*

This is well known. For $0 < s \in S_n(\mathbf{K})$ we write $s^{1/2} = r$ with the above element r ; we then put $s^{-1/2} = (s^{-1})^{1/2}$.

Now the higher-dimensional analogue of (II) of the introduction is given by

Theorem 2. *Given $x \in M_n^m(\mathbf{K})$, there exist $c \in M_n^m(\mathbf{K})$ and $d \in GL_m(\mathbf{K})$ such that $x = d^{-1}c$ and $dd^* + cc^* = 1_m$. If $x = d'^{-1}c'$ is another such expression, then $c' = hc$ and $d' = hd$ with $h \in U_m(\mathbf{K})$. Moreover there is a unique choice of such $(c \ d)$ for x with the property that $0 < d \in S_m(\mathbf{K})$.*

Proof: Observing that $1_m + xx^*$ is positive definite, we find an element e of $GL_m(\mathbf{K})$ such that $1_m + xx^* = ee^*$. Putting $d = e^{-1}$ and $c = dx$, we obtain the first assertion. Given another such pair (c', d') , put $h = d'e$. Then $d' = hd$, $c' = hc$, and $hh^* = d'ee^*d'^* = d'(1_m + xx^*)d'^* = d'd'^* + c'c'^* = 1_m$, and hence $h \in U_m(\mathbf{K})$. By Lemma 3 we can put $e = (1_m + xx^*)^{1/2}$. Then $0 < d \in S_m(\mathbf{K})$. If $0 < d' \in S_m(\mathbf{K})$, then $d'^2 = d^*h^*hd = d^2$, and hence $d' = d$ by Lemma 3. This completes the proof.

Remarks. (R4) In the above theorem, we can take d to be an upper triangular matrix with positive diagonal elements. When d is so chosen, the pair (c, d) is unique for x . This can be shown as in the above proof by means of Lemma 3. Can we prove a similar uniqueness in Theorem 1? We shall answer this question in (R7) below.

(R5) We are tempted to call an element X of $M_m^m(\mathbf{K})$, $m \leq n$, *primitive* if $XX^* = 1_m$, though that may not be good terminology. If we do use it, then $(c \ d)$ with c, d as in Theorem 2 is primitive. Moreover, the analogue of (1) \Leftrightarrow (2) of Lemma 2 can be stated as follows:

An element X of $M_n^m(\mathbf{K})$, $m \leq n$, is primitive if and only if there is an element Y of $M_n^{n-m}(\mathbf{K})$ such that $\begin{pmatrix} X \\ Y \end{pmatrix} \in U_n(\mathbf{K})$.

Now the last part of Theorem 2 and its proof can be stated in the following way.

Lemma 4. Put $\varphi(x) = (1_m + xx^*)^{-1/2}x$ for $x \in M_n^m(\mathbf{K})$. Then φ gives a one-to-one map of $M_n^m(\mathbf{K})$ onto $\{y \in M_n^m(\mathbf{K}) | 1_m > yy^*\}$, and the inverse map ψ of φ is given by $\psi(y) = (1_m - yy^*)^{-1/2}y$.

This is an easy exercise (cf. [2, Lemma 2.3]).

3. A GROUP-THEORETICAL INTERPRETATION. Let us now show that there are some group-theoretical facts which give more substance to the parallelism between Theorems 1 and 2. For $F = \mathbf{Q}$ or \mathbf{K} (or for any ring F) we define subgroups $P_n(F)$ of $GL_n(F)$ and $P_{n,m}(F)$ of $GL_{m+n}(F)$ by

$$P_{n,m}(F) = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in GL_{m+n}(F) \mid a \in GL_n(F), b \in M_m^n(F), d \in GL_m(F) \right\},$$

$$P_n(F) = \bigcap_{r=0}^n P_{r,n-r}(F),$$

where we understand that $P_{n,0}(F) = P_{0,n}(F) = GL_n(F)$. Clearly $P_n(F)$ is the group of all upper triangular matrices of $GL_n(F)$.

Theorem 3. (1) $GL_n(\mathbf{Q}) = P_n(\mathbf{Q})GL_n(\mathbf{Z}) = P_{r,n-r}(\mathbf{Q})GL_n(\mathbf{Z})$ ($0 \leq r \leq n$).

(2) $GL_n(\mathbf{K}) = P_n(\mathbf{K})U_n(\mathbf{K}) = P_{r,n-r}(\mathbf{K})U_n(\mathbf{K})$ ($0 \leq r \leq n$).

Proof: Though these are well known, we give a proof here for the reader's convenience. Since $P_{r,n-r}(F) \subset P_n(F)$, it is sufficient to prove the first equality in each case. We first prove Case (1) by induction on n . It is trivial if $n = 1$. Given $\xi \in GL_n(\mathbf{Q})$, $n > 1$, let x be the last row of ξ . Then $x = qy$ with $0 \neq q \in \mathbf{Q}$ and a primitive element y of $M_n^1(\mathbf{Z})$. Take an element α of $GL_n(\mathbf{Z})$ whose last row is y . Then $y = (0_{n-1}^1 \ 1)\alpha$, so that $x\alpha^{-1} = (0_{n-1}^1 \ q)$. Thus we can put $\xi\alpha^{-1} = \begin{pmatrix} r & s \\ 0 & q \end{pmatrix}$ with $r \in M_{n-1}^{n-1}(\mathbf{Q})$ and $s \in M_{1,n-1}^1(\mathbf{Q})$. By induction we find $\tau \in P_{n-1}(\mathbf{Q})$ and $\sigma \in GL_{n-1}(\mathbf{Z})$ such that $r = \tau\sigma$. Then $\xi\alpha^{-1} \cdot \text{diag}[\sigma^{-1}, 1] \in P_n(\mathbf{Q})$, which gives (1). To prove (2), let $\xi \in GL_n(\mathbf{K})$. By Lemma 3 we can find $\eta \in P_n(\mathbf{K})$ such that $\xi\xi^* = \eta\eta^*$. Then $\eta^{-1}\xi \in U_n(\mathbf{K})$ and $\xi = \eta \cdot \eta^{-1}\xi$, which proves (2).

The groups $P_{n,m}$ and P_n are examples of *parabolic subgroups* of algebraic groups, and the decompositions of the above types are well known for classical groups. We shall give some more examples in Theorem 5 below.

Let us now derive the essential part of Theorems 1 and 2 from Theorem 3. Given $x \in M_n^m(\mathbf{Q})$, we consider an element

$$\begin{pmatrix} 1_n & 0_m^n \\ x & 1_m \end{pmatrix} \text{ of } GL_{m+n}(\mathbf{Q}).$$

By (1) of Theorem 3 we have

$$\begin{pmatrix} 1_n & 0_m^n \\ x & 1_m \end{pmatrix} = \begin{pmatrix} p & q \\ 0 & s \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (3.1)$$

with

$$\begin{pmatrix} p & q \\ 0 & s \end{pmatrix} \in P_{n,m}(\mathbf{Q}) \quad \text{and} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_{n+m}(\mathbf{Z}).$$

Then $\begin{pmatrix} c & d \end{pmatrix}$ is primitive and $\begin{pmatrix} x & 1 \\ sc & sd \end{pmatrix}$. Therefore d is invertible and $x = d^{-1}c$.

Similarly, for $x \in M_n^m(\mathbf{K})$ we have (3.1) with

$$\begin{pmatrix} p & q \\ 0 & s \end{pmatrix} \in P_{n,m}(\mathbf{K}) \quad \text{and} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in U_{m+n}(\mathbf{K}).$$

Again d is invertible and $x = d^{-1}c$. Since $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in U_{m+n}(\mathbf{K})$, we have $cc^* + dd^* = 1$ (or $\begin{pmatrix} c & d \end{pmatrix}$ is primitive in the sense of (R5)).

Thus the parallelism between Theorems 1 and 2 can be interpreted as the parallelism between (1) and (2) of Theorem 3, or rather, between the decompositions given by (3.1) in both cases.

Theorem 3 positions $U_n(\mathbf{K})$ as the counterpart of $GL_n(\mathbf{Z})$. As the counterpart of Lemma 1 in this sense we obtain

Lemma 5. *Given $X \in M_n^m(\mathbf{K})$ of rank r , there exist $A \in U_m(\mathbf{K})$, $B \in U_n(\mathbf{K})$ and positive real numbers e_1, \dots, e_r such that $e_i \leq e_{i+1}$ for all $i < r$ and (2.1) holds. Moreover, the e_i are uniquely determined by X .*

The statement for X belonging to the semisimple part of $GL_n(\mathbf{K})$ is a special case of the fact in the theory of Lie groups which is usually stated as $G = K\bar{A}_+K$ (cf. [1, p. 402]).

Proof: The uniqueness follows from the easy fact that the e_i^2 are the nonzero eigenvalues of XX^* . To obtain the desired expression, we view (left matrix multiplication by) X as a (right-) \mathbf{K} -linear map of $M_1^n(\mathbf{K})$ into $M_1^m(\mathbf{K})$. Let V be the orthogonal complement of $\text{Ker}(X)$ in $M_1^n(\mathbf{K})$. Then X is essentially a map of V onto XV . This reduces our problem to the case in which $m = n$ and $X \in GL_n(\mathbf{K})$. In this case we can find an element A of $U_n(\mathbf{K})$ so that $AXX^*A^* = \text{diag}[f_1, \dots, f_n]$, $0 < f_1 < \dots < f_n$. Let $e_i = f_i^{1/2}$ and $B = X^{-1}A^{-1} \text{diag}[e_1, \dots, e_n]$. Then $BB^* = 1_n$ and we obtain the desired expression.

Remarks. (R6) As mentioned in (R3), we can prove the same types of results with any p -adic field F and its ring A of p -adic integers in place of \mathbf{Q} and \mathbf{Z} . In this case, $GL_n(A)$ is a maximal compact subgroup of $GL_n(F)$, as $U_n(\mathbf{K})$ is in $GL_n(\mathbf{K})$. (The fact follows easily from the p -adic version of Lemma 1 and Lemma 5.) This kind of analogy belongs to the standard philosophy in number theory today, but we shall not make any p -adic discussion in this article.

(R7) In Theorem 1, for a given x the element d can be changed for any element in the coset $GL_m(\mathbf{Z})d$. This means that if we fix a complete set of representatives D for $GL_m(\mathbf{Z}) \backslash GL_m(\mathbf{Q})$, we can choose d from D . Now by Theorem 3(1) we have $GL_m(\mathbf{Q}) = GL_m(\mathbf{Z})P_m(\mathbf{Q})$. Therefore we can take D to be a subset of $P_m(\mathbf{Q})$. More precisely, if we take a complete set of representatives D_0 for $[GL_m(\mathbf{Z}) \cap P_m(\mathbf{Q})] \backslash P_m(\mathbf{Q})$, then, given x , there exists a unique $\begin{pmatrix} c & d \end{pmatrix}$ as in Theorem 1(1) with $d \in D_0$.

4. FRACTIONAL EXPRESSIONS FOR HERMITIAN AND SKEWHERMITIAN MATRICES. By restricting our matrices to symmetric, alternating, hermitian, or skewhermitian ones, we can still make clear-cut statements on their fractional expressions. Though the results follow immediately from Theorems 1 and 2, they reflect the decompositions of symplectic, orthogonal, and unitary groups in the same sense as in Theorem 3, and therefore, may make the parallelism more cogent.

With A denoting \mathbf{Z} , \mathbf{Q} , or \mathbf{K} and the symbol $\varepsilon = (\varepsilon_1, \varepsilon_2)$ with $\varepsilon_\nu = +$ or $-$ we put

$$S^\varepsilon(A) = \{x \in M_n^n(A) | x^\rho = -\varepsilon_1 x\},$$

$$G^\varepsilon(A) = \{X \in GL_{2n}(A) | XJ_\varepsilon X^\rho = J_\varepsilon\}, \quad J_\varepsilon = \begin{pmatrix} 0 & \varepsilon_1 1_n \\ 1_n & 0 \end{pmatrix},$$

$$C^\varepsilon(\mathbf{K}) = G^\varepsilon(\mathbf{K}) \cap U_{2n}(\mathbf{K}),$$

$$V^\varepsilon(A) = \{w \in M_{2n}^n(A) | wJ_\varepsilon w^\rho = 0, \text{rank}(w) = n\}.$$

Here $X^\rho = X^*$ if $\varepsilon_2 = +$ and $X^\rho = {}^t X$ if $\varepsilon_2 = -$; we consider $\varepsilon_2 = -$ only when $A = \mathbf{K} = \mathbf{C}$. If $A = \mathbf{H}$, the condition $\text{rank}(w) = n$ means that ${}_w M_1^{2n}(\mathbf{H}) = M_1^n(\mathbf{H})$. Clearly $S^{(-,+)}(\mathbf{K}) = S_n(\mathbf{K})$. If $w = \begin{pmatrix} c & d \end{pmatrix} \in M_{2n}^n(A)$ with $c, d \in M_n^n(A)$ and $\text{rank}(w) = n$, then

$$w \in V^\varepsilon(A) \Leftrightarrow cd^\rho = -\varepsilon_1 dc^\rho \Leftrightarrow cd^\rho \in S^\varepsilon(A). \quad (4.1)$$

If A is \mathbf{Z} , \mathbf{Q} , or \mathbf{R} , the group $G^{(-,+)}(A)$ is usually denoted by $Sp(n, A)$. If $w \in V^\varepsilon(A)$, $\alpha \in GL_n(A)$, and $\beta \in G^\varepsilon(A)$, then $\alpha w \beta \in V^\varepsilon(A)$. Since $\begin{pmatrix} 0 & 1_n \end{pmatrix} \in V^\varepsilon(A)$, we have $\begin{pmatrix} 0 & 1_n \end{pmatrix} \beta \in V^\varepsilon(A)$. Clearly $\begin{pmatrix} 0 & 1_n \end{pmatrix} \beta$ is the lower half of β . We also put

$$W^\varepsilon(\mathbf{Z}) = \{w \in V^\varepsilon(\mathbf{Z}) | w \text{ is primitive}\}, \quad (4.2)$$

$$W^\varepsilon(\mathbf{K}) = \{w \in V^\varepsilon(\mathbf{K}) | ww^* = 1_n\}. \quad (4.3)$$

Clearly an element $\begin{pmatrix} c & d \end{pmatrix}$ of $V^\varepsilon(\mathbf{K})$ belongs to $W^\varepsilon(\mathbf{K})$ if and only if $cc^* + dd^* = 1_n$. Now our fractional expressions for the matrices in $S^\varepsilon(A)$ are given by

Theorem 4. (1) Given $x \in S^\varepsilon(\mathbf{Q})$, there exists $\begin{pmatrix} c & d \end{pmatrix} \in W^\varepsilon(\mathbf{Z})$ with an invertible d such that $x = d^{-1}c$.

(2) Given $x \in S^\varepsilon(\mathbf{K})$, there exists $\begin{pmatrix} c & d \end{pmatrix} \in W^\varepsilon(\mathbf{K})$ with an invertible d such that $x = d^{-1}c$.

Proof: For $x \in S^\varepsilon(\mathbf{Q})$ take c and d as in (1) of Theorem 1. Then ${}^t c \cdot {}^t d^{-1} = {}^t(d^{-1}c) = -\varepsilon_1 d^{-1}c$, and hence $\begin{pmatrix} c & d \end{pmatrix} \in W^\varepsilon(\mathbf{Z})$, which proves (1). Assertion (2) can be derived from Theorem 2 in the same fashion.

As the analogue of Lemma 2(2) and the statement in (R5) we obtain:

Lemma 6. (1) $W^\varepsilon(\mathbf{Z}) = \begin{pmatrix} 0 & 1_n \end{pmatrix} G^\varepsilon(\mathbf{Z})$.

(2) $W^\varepsilon(\mathbf{K}) = \begin{pmatrix} 0 & 1_n \end{pmatrix} C^\varepsilon(\mathbf{K})$. More precisely, $\alpha \mapsto \begin{pmatrix} 0 & 1_n \end{pmatrix} \alpha$ gives a one-to-one map of $C^\varepsilon(\mathbf{K})$ onto $W^\varepsilon(\mathbf{K})$.

Proof: To prove (1), we first observe that $\begin{pmatrix} 0 & 1_n \end{pmatrix} G^\varepsilon(\mathbf{Z}) \subset W^\varepsilon(\mathbf{Z})$ since $\begin{pmatrix} 0 & 1_n \end{pmatrix}$ is primitive and $G^\varepsilon(\mathbf{Z}) \subset GL_{2n}(\mathbf{Z})$. Conversely, take $x \in W^\varepsilon(\mathbf{Z})$. By (2) of Lemma 2,

we can find some $y \in \mathbf{Z}_{2n}^n$ such that $\begin{pmatrix} y \\ x \end{pmatrix} \in GL_{2n}(\mathbf{Z})$. Put $\alpha = \begin{pmatrix} y \\ x \end{pmatrix}$. Then

$$\alpha J_\varepsilon \cdot {}^t \alpha = \begin{pmatrix} y \\ x \end{pmatrix} (J_\varepsilon \cdot {}^t y \quad J_\varepsilon \cdot {}^t x) = \begin{pmatrix} u & v \\ \varepsilon_1 \cdot {}^t v & 0 \end{pmatrix}$$

with $u, v \in M_n^n(\mathbf{Z})$. Since $\alpha J_\varepsilon \cdot {}^t \alpha \in GL_{2n}(\mathbf{Z})$, we see that $v \in GL_n(\mathbf{Z})$. Put $\beta = \text{diag}[\varepsilon_1 v^{-1}, 1_n]$. Then

$$\beta \alpha J_\varepsilon \cdot {}^t \alpha \cdot {}^t \beta = \begin{pmatrix} z & \varepsilon_1 1_n \\ 1_n & 0 \end{pmatrix}$$

with $z \in M_n^n(\mathbf{Z})$. If $(a \ b)$ is the upper half of $\beta \alpha$, then $z = \varepsilon_1 a \cdot {}^t b + b \cdot {}^t a$. Put

$$\gamma = \begin{pmatrix} 1_n & -b \cdot {}^t a \\ 0 & 1_n \end{pmatrix}.$$

Then $\gamma \beta \alpha J_\varepsilon \cdot {}^t \alpha \cdot {}^t \beta \cdot {}^t \gamma = J_\varepsilon$, and so $\gamma \beta \alpha \in G^\varepsilon(\mathbf{Z})$. Now we see that $(0 \ 1_n) \gamma \beta = (0 \ 1_n)$, and hence $(0 \ 1_n) \gamma \beta \alpha = (0 \ 1_n) \alpha = x$, which proves (1). Assertion (2) follows from

$$C^\varepsilon(\mathbf{K}) = \left\{ \begin{pmatrix} d' & \varepsilon_1 c' \\ c & d \end{pmatrix} \middle| (c \ d) \in W^\varepsilon(\mathbf{K}) \right\}, \quad (4.4)$$

where $(d' \ c') = (\bar{d} \ \bar{c})$ if $\mathbf{K} = \mathbf{C}$ and $\varepsilon_2 = -$; $(d' \ c') = (d \ c)$ otherwise. In fact, it is easy to verify that any matrix on the right-hand side is contained in $C^\varepsilon(\mathbf{K})$. Conversely, let

$$\alpha \in C^\varepsilon(\mathbf{K}), \quad (0 \ 1_n) \alpha = (c \ d), \quad \text{and} \quad \beta = \begin{pmatrix} d' & \varepsilon_1 c' \\ c & d \end{pmatrix}.$$

Then $\beta \in C^\varepsilon(\mathbf{K})$ and $(0 \ 1_n) \alpha = (0 \ 1_n) \beta$, and hence $\alpha \beta^{-1} = \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$ with $a, b \in M_n^n(\mathbf{K})$. Since $\alpha \beta^{-1} \in C^\varepsilon(\mathbf{K})$, we see that $a = 1$ and $b = 0$, which proves (4.4).

To obtain the analogue of Theorem 3 in the present case, put

$$P^\varepsilon(\mathbf{Q}) = G^\varepsilon(\mathbf{Q}) \cap P_{n,n}(\mathbf{Q}), \quad P^\varepsilon(\mathbf{K}) = G^\varepsilon(\mathbf{K}) \cap P_{n,n}(\mathbf{K}).$$

Theorem 5. (1) $G^\varepsilon(\mathbf{Q}) = P^\varepsilon(\mathbf{Q}) G^\varepsilon(\mathbf{Z})$.

(2) $G^\varepsilon(\mathbf{K}) = P^\varepsilon(\mathbf{K}) C^\varepsilon(\mathbf{K})$.

Proof: To prove (1), let $\xi \in G^\varepsilon(\mathbf{Q})$. By Theorem 3(1) we have $\xi = \eta \alpha$ with $\eta \in P_{n,n}(\mathbf{Q})$ and $\alpha \in GL_{2n}(\mathbf{Z})$. Put

$$\eta = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \quad \text{and} \quad \xi = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$$

with a, b , etc. of size n . Then $d^{-1}(r \ s) = (0 \ 1_n) \alpha$. Since $\xi \in G^\varepsilon(\mathbf{Q})$, the left-hand side of the last equality, $d^{-1}(r \ s)$, belongs to $V^\varepsilon(\mathbf{Q})$. The right-hand side $(0 \ 1_n) \alpha$ is primitive. Thus $(0 \ 1_n) \alpha \in W^\varepsilon(\mathbf{Z})$. By Lemma 6(1) we have $(0 \ 1_n) \alpha = (0 \ 1_n) \beta$ with $\beta \in G^\varepsilon(\mathbf{Z})$. Put $\gamma = \alpha \beta^{-1}$. Then $(0 \ 1_n) \gamma = (0 \ 1_n)$, and hence $\gamma \in P_{n,n}(\mathbf{Q})$. Now $\xi = \eta \alpha = \eta \gamma \beta$, which proves (1) if we can show that $\eta \gamma \in P^\varepsilon(\mathbf{Q})$, but this is indeed the case, since $\eta \gamma \in P_{n,n}(\mathbf{Q})$ and $\eta \gamma = \xi \beta^{-1} \in G^\varepsilon(\mathbf{Q})$. Similarly let $\xi \in G^\varepsilon(\mathbf{K})$. By Theorem 3(2) we have $\xi = \eta \alpha$ with $\eta \in P_{n,n}(\mathbf{K})$ and $\alpha \in U_{2n}(\mathbf{K})$. With d, r , and s as above, $(0 \ 1_n) \alpha = d^{-1}(r \ s) \in W^\varepsilon(\mathbf{K})$. By Lemma 6(2) we have $(0 \ 1_n) \alpha = (0 \ 1_n) \beta$ with $\beta \in C^\varepsilon(\mathbf{K})$. Repeating the above argument with \mathbf{K} in place of \mathbf{Q} , we obtain (2).

The above theorem enables us to explain Theorem 4 again by means of the decomposition of (3.1). Indeed, if $x \in S^\varepsilon(A)$, then $\begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix} \in G^\varepsilon(A)$. By Theorem 5 we have (3.1) with $\begin{pmatrix} p & q \\ 0 & s \end{pmatrix} \in P^\varepsilon(A)$ and $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in C$, where C denotes $C^\varepsilon(\mathbf{K})$ or $G^\varepsilon(\mathbf{Z})$ according as $A = \mathbf{K}$ or \mathbf{Q} . Then we obtain $x = d^{-1}c$ with $(c \ d) \in W^\varepsilon(\mathbf{K})$ or $W^\varepsilon(\mathbf{Z})$.

We conclude this section by adding some historical notes. The expression $x = d^{-1}c$ of a symmetric rational matrix x by means of $(c \ d)$ belonging to $W^{(\cdot, +)}(\mathbf{Z})$ was first considered by Siegel in [3, p. 653]. In particular, he observes that $|\det(d)|$ is the product of the reduced denominators of the elementary divisors of x (as shown in Theorem 1(3)), and puts $\nu(x) = |\det(d)|$. He employed this function ν in his investigations of Eisenstein series in [3] and of indefinite quadratic forms in [4]. It should also be noted that Lemma 6(1) for $G^\varepsilon(\mathbf{Q}) = Sp(n, \mathbf{Q})$ was first given by Siegel. The decomposition of an algebraic group over a local field as in Theorems 3 and 5, as well as the generalization of Siegel's function ν , has been employed in recent papers on automorphic forms.

5. MATRIX-VALUED TRIGONOMETRIC FUNCTIONS. As mentioned at the beginning, c and d in the one-dimensional case can be given as $\cos \theta$ and $\sin \theta$. Let us now give higher-dimensional analogues of these trigonometric functions which play similar roles. With fixed m and n we put

$$W(\mathbf{K}) = \{(c \ d) \in M_n^m(\mathbf{K}) \times M_m^m(\mathbf{K}) | cc^* + dd^* = 1_m\},$$

$$W'(\mathbf{K}) = \{(c \ d) \in W(\mathbf{K}) | d = d^*\}.$$

For $X \in M_n^m(\mathbf{K})$ we put $\exp(X) = \sum_{\nu=0}^{\infty} X^\nu / \nu!$ as usual, and we define an $M_n^m(\mathbf{K})$ -valued function \mathbf{c} , an $M_m^m(\mathbf{K})$ -valued function \mathbf{d} , and a $GL_{m+n}(\mathbf{K})$ -valued function E on $M_n^m(\mathbf{K})$ by

$$\mathbf{c}(y) = \sum_{\nu=0}^{\infty} (-yy^*)^\nu y / (2\nu + 1)!, \quad \mathbf{d}(y) = \sum_{\nu=0}^{\infty} (-yy^*)^\nu / (2\nu)!,$$

$$E(y) = \exp \left(\begin{pmatrix} 0_n & -y^* \\ y & 0_m^m \end{pmatrix} \right) \quad (y \in M_n^m(\mathbf{K})).$$

We have clearly

$$\mathbf{c}(y)^* = \mathbf{c}(y^*), \quad \mathbf{d}(y)^* = \mathbf{d}(y), \quad {}^t\mathbf{c}(y) = \mathbf{c}({}^ty), \quad {}^t\mathbf{d}(y) = \mathbf{d}(\bar{y}),$$

$$E(y) = \begin{pmatrix} \mathbf{d}(y^*) & -\mathbf{c}(y)^* \\ \mathbf{c}(y) & \mathbf{d}(y) \end{pmatrix}.$$

Since $\exp(X) \in U_N(\mathbf{K})$ if $X^* = -X \in M_N^N(\mathbf{K})$, the matrix $E(y)$ belongs to $U_{m+n}(\mathbf{K})$, and hence

$$\mathbf{c}(y)\mathbf{c}(y)^* + \mathbf{d}(y)^2 = 1_m, \quad \mathbf{c}(y)^*\mathbf{c}(y) + \mathbf{d}(y^*)^2 = 1_n, \\ \mathbf{c}(y)\mathbf{d}(y^*) = \mathbf{d}(y)\mathbf{c}(y).$$

Theorem 6. Every element x of $M_n^m(\mathbf{K})$ can be written $x = \mathbf{d}(y)^{-1}\mathbf{c}(y)$ with an element $y \in M_n^m(\mathbf{K})$ such that $\mathbf{d}(y)$ is invertible. Moreover, we have

$$W(\mathbf{K}) = U_m(\mathbf{K})W'(\mathbf{K}) = U_m(\mathbf{K})\{(\mathbf{c}(y) \ \mathbf{d}(y)) | y \in M_n^m(\mathbf{K})\}, \quad (5.1)$$

$$W'(\mathbf{K}) = \{(\mathbf{c}(y) \ \mathbf{d}(y)) | y \in M_n^m(\mathbf{K})\} \quad \text{if } m \leq n, \quad (5.2)$$

$$U_{m+n}(\mathbf{K}) = (U_n(\mathbf{K}) \times U_m(\mathbf{K}))E(M_n^m(\mathbf{K})), \quad (5.3)$$

where $U_n(\mathbf{K}) \times U_m(\mathbf{K}) = \{\text{diag}[a, b] | a \in U_n(\mathbf{K}), b \in U_m(\mathbf{K})\}$.

Proof: The first assertion follows from Theorem 2 and (5.1). Let $(c \ d) \in W(\mathbf{K})$. By Lemma 5 we have $d = uev$ with $u, v \in U_m(\mathbf{K})$ and a real diagonal matrix e . Put $w = v^{-1}u^{-1}$ and $(e \ f) = w(c \ d)$. Then $(e \ f) \in W(\mathbf{K})$ and $f = v^{-1}ev$. Thus $f^* = f$ and so $(e \ f) \in W'(\mathbf{K})$, which proves the first equality of (5.1). Assuming now $m \leq n$, let us prove (5.2). Clearly the right-hand side is contained in the left-hand side. To prove the opposite inclusion, let $(c \ d) \in W'(\mathbf{K})$. Take $p \in U_m(\mathbf{K})$ so that $pd p^{-1} = \text{diag}[a, -1_s, 1_t]$, $m = r + s + t$, with a real diagonal matrix a of size r whose diagonal entries are different from ± 1 . Then $pc(pc)^* = 1_r - pd^2 p^{-1} = \text{diag}[1_r - a^2, 0_{m-r}^r]$. From this we can conclude that

$$pc = \begin{pmatrix} z \\ 0_{n-r}^{m-r} \end{pmatrix} \quad \text{with } z \in M_n^r(\mathbf{K}).$$

We can then find a real diagonal matrix b of size r such that $b^2 = 1_r - a^2$, and also a diagonal matrix σ of size r such that $\exp(i\sigma) = ib + a$. By Lemma 5 we can find $q \in U_n(\mathbf{K})$ such that $z = (v \ 0_{n-r}^r)q$ with $v \in GL_r(\mathbf{K})$. Then $vv^* = zz^* = b^2$. Put

$$y = p^{-1} \begin{pmatrix} \sigma b^{-1}v & 0_s^r & 0_u^r \\ 0_r^s & \pi 1_s & 0_u^s \\ 0_r^t & 0_s^t & 0_u^t \end{pmatrix} q,$$

where $u = n - r - s$ and π is a real number such that $e^{i\pi} = -1$. Then $yy^* = p^{-1} \text{diag}[\sigma^2, \pi^2 1_s, 0]p$, and a straightforward calculation shows that $\mathbf{c}(y) = c$ and $\mathbf{d}(y) = d$, which proves (5.2), and the second equality of (5.1) as well, under the condition $m \leq n$. If $m > n$, this reasoning fails since $r + s$ may be greater than n . To avoid this difficulty, after taking p, a, b , and σ for a given $(c \ d) \in W'(\mathbf{K})$, we put $u = p^{-1} \text{diag}[1_r, -1_s, 1_t]p$ and $(e \ f) = u(c \ d)$. Then $pfp^{-1} = \text{diag}[a, 1_{s+t}]$, $(pe)(pe)^* = \text{diag}[b^2, 0]$, and

$$pe = \begin{pmatrix} z \\ 0_{n-r}^{m-r} \end{pmatrix} \quad \text{with } z \in M_n^r(\mathbf{K}).$$

Taking v and q as above, put

$$y = p^{-1} \begin{pmatrix} \sigma b^{-1}v & 0_{n-r}^r \\ 0_{n-r}^{m-r} & 0_{n-r}^{m-r} \end{pmatrix} q.$$

Then $yy^* = p^{-1} \text{diag}[\sigma^2, 0]p$, $\mathbf{c}(y) = e$, and $\mathbf{d}(y) = f$, which proves (5.1) in general. To prove (5.3), let $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in U_{m+n}(\mathbf{K})$. Since $(c \ d) \in W(\mathbf{K})$, we have $(c \ d) = u(\mathbf{c}(y) \ \mathbf{d}(y))$ with $u \in U_m(\mathbf{K})$ and $y \in M_n^m(\mathbf{K})$ by (5.1). Put $\beta = E(y)$. Then $(0 \ 1_m)\alpha = (0 \ u)\beta$, and hence we can put $\text{diag}[1_n, u^{-1}]\alpha\beta^{-1} = \begin{pmatrix} g & h \\ 0 & 1_n \end{pmatrix}$ with $g, h \in M_n^n(\mathbf{K})$. Since the left-hand side belongs to $U_{m+n}(\mathbf{K})$, we see that $g \in U_n(\mathbf{K})$ and $h = 0$. This proves (5.3), since the opposite inclusion is obvious.

It should be noted that the equality of (5.2) is false if $m > n$. In fact, take $(c \ d) = (0 \ -1_m)$. Suppose that $m > n$ and $d = \mathbf{d}(y)$ with $y \in M_n^m(\mathbf{K})$. We can find $p \in U_m(\mathbf{K})$ and a real diagonal matrix r such that $yy^* = pr^2 p^{-1}$. Then $-1_m = p^{-1}\mathbf{d}(y)p = \mathbf{d}(r)$, and hence $r = \text{diag}[a_1\pi, \dots, a_m\pi]$ with odd integers a_ν . This is a contradiction, since $\text{rank}(yy^*) \leq n < m$.

To consider the case of matrices in $S^e(\mathbf{K})$, we put

$$Z^e(\mathbf{K}) = \{(c \ d) \in W^e(\mathbf{K}) | c^\rho = -\varepsilon_1 c, d^* = d\}$$

and restrict the functions \mathbf{c} and \mathbf{d} to $S^\varepsilon(\mathbf{K})$. Clearly, for $s \in S^\varepsilon(\mathbf{K})$ we have

$$\begin{aligned} \mathbf{c}(s)^\rho &= -\varepsilon_1 \mathbf{c}(s), & \mathbf{d}(s)^* &= \mathbf{d}(s), & \mathbf{d}(s)^\rho &= \mathbf{d}(s^*), \\ \mathbf{c}(s)\mathbf{d}(s)^\rho &= \mathbf{d}(s)\mathbf{c}(s), & \mathbf{c}(s)\mathbf{c}(s)^* + \mathbf{d}(s)^2 &= 1_n. \end{aligned}$$

From these and (4.4) we see that

$$\begin{aligned} \{(\mathbf{c}(s) \quad \mathbf{d}(s)) | s \in S^\varepsilon(\mathbf{K})\} &\subset Z^\varepsilon(\mathbf{K}) \\ = \{(c \quad d) \in M_{2n}^n(\mathbf{K}) | c^\rho &= -\varepsilon_1 c, d^* = d, cd^\rho = dc, cc^* + dd^* = 1_n\}, \end{aligned} \quad (5.4)$$

$$E(S^\varepsilon(\mathbf{K})) \subset \{\alpha \in C^\varepsilon(\mathbf{K}) | \alpha^* \lambda = \lambda \alpha\}, \quad \lambda = \text{diag}[1_n, -1_n]. \quad (5.5)$$

Theorem 7. *Every element x of $S^\varepsilon(\mathbf{K})$ can be written $x = \mathbf{d}(s)^{-1}\mathbf{c}(s)$ with an element $s \in S^\varepsilon(\mathbf{K})$ such that $\mathbf{d}(s)$ is invertible. Moreover, excluding the two cases in which $(\varepsilon, \mathbf{K}) = ((+, +), \mathbf{R})$ or $((+, -), \mathbf{C})$, we have*

$$W^\varepsilon(\mathbf{K}) = U_n(\mathbf{K})Z^\varepsilon(\mathbf{K}), \quad (5.6)$$

$$Z^\varepsilon(\mathbf{K}) = \{(\mathbf{c}(s) \quad \mathbf{d}(s)) | s \in S^\varepsilon(\mathbf{K})\}, \quad (5.7)$$

$$E(S^\varepsilon(\mathbf{K})) = \{\alpha \in C^\varepsilon(\mathbf{K}) | \alpha^* \lambda = \lambda \alpha\}, \quad (5.8)$$

$$C^\varepsilon(\mathbf{K}) = \{\text{diag}[u', u] | u \in U_n(\mathbf{K})\}E(S^\varepsilon(\mathbf{K})), \quad (5.9)$$

where $u' = \bar{u}$ if $(\varepsilon, \mathbf{K}) = ((-, -), \mathbf{C})$ and $u' = u$ otherwise.

Proof: The first assertion follows from Theorem 4(2), (5.6), and (5.7) except in the two exceptional cases, which will be treated at the end of the proof. To prove (5.7), let $(c \quad d) \in Z^\varepsilon(\mathbf{K})$. We repeat the proof of Theorem 6 with some modifications. Taking p and a as in that proof, we put $g = pcp^\rho$ and $h = pdp^{-1}$. Then $g^\rho = -\varepsilon_1 g$, $gg^* = \text{diag}[1 - a^2, 0]$, and $gh^\rho = hg$. From this we can conclude that $g = \text{diag}[v, 0]$ with $v = -\varepsilon_1 v^\rho \in M_r^r(\mathbf{K})$, $va = av$. We may assume that $a = \text{diag}[a_1 1_{r_1}, \dots, a_k 1_{r_k}]$ with $r = r_1 + \dots + r_k$ and the a_ν which are all different. Then $v = \text{diag}[v_1, \dots, v_k]$ with $v_\nu = -\varepsilon_1 v_\nu^\rho \in M_{r_\nu}^{r_\nu}(\mathbf{K})$. Put $b = \text{diag}[b_1 1_{r_1}, \dots, b_k 1_{r_k}]$ and $\sigma = \text{diag}[\sigma_1 1_{r_1}, \dots, \sigma_k 1_{r_k}]$ with real numbers b_ν and σ_ν such that $b_\nu^2 = 1 - a_\nu^2$ and $\exp(i\sigma_\nu) = ib_\nu + a_\nu$. If $\varepsilon_1 = -$, put $y = p^{-1} \text{diag}[\sigma b^{-1} v, \pi 1_s, 0](p^\rho)^{-1}$. Then $y \in S^\varepsilon(\mathbf{K})$, $c = \mathbf{c}(y)$, and $d = \mathbf{d}(y)$. If $\varepsilon = (+, +)$ and $\mathbf{K} \neq \mathbf{R}$, then we get the same conclusion with $y = p^{-1} \text{diag}[\sigma b^{-1} v, i\pi 1_s, 0]p$. This combined with (5.4) proves (5.7). Then (5.8) follows immediately from (5.7) and (4.4). To prove (5.9), given $\alpha \in C^\varepsilon(\mathbf{K})$, put $\beta = \lambda \alpha^* \lambda \alpha$. Then $\beta \in C^\varepsilon(\mathbf{K})$ and $\beta^* = \lambda \beta \lambda$, so that $\beta = E(y)$ with $y \in S^\varepsilon(\mathbf{K})$ by (5.7). Put $\gamma = E(-y/2)$. Then $\lambda \alpha^{-1} \lambda \alpha = \beta = \gamma^2 = \lambda \gamma^{-1} \lambda \gamma$, and hence $\lambda \gamma \alpha^{-1} \lambda = \gamma \alpha^{-1}$. This shows that $\gamma \alpha^{-1}$ commutes with λ , which implies that $\gamma \alpha^{-1} = \text{diag}[w, z]$ with $w, z \in U_n(\mathbf{K})$. Since $\gamma \alpha^{-1} \in C^\varepsilon(\mathbf{K})$, we see that $w = z'$. This proves (5.9). Finally, (5.6) follows from Lemma 6(2), (5.9), and (5.4).

In the exceptional cases we proceed as follows: Put $u = p^{-1} \text{diag}[1_r, -1_s, 1_t]p$ and $(e \quad f) = u(c \quad d)$. Then $pep^\rho = g$, $pfp^{-1} = \text{diag}[a, 1_{s+t}]$. Putting $y = p^{-1} \text{diag}[\sigma b^{-1} v, 0](p^\rho)^{-1}$, we obtain $y \in S^\varepsilon(\mathbf{K})$, $e = \mathbf{c}(y)$, and $f = \mathbf{d}(y)$. Thus, in the exceptional cases we have

$$Z^\varepsilon(\mathbf{K}) \subset U_n(\mathbf{K})\{(\mathbf{c}(s) \quad \mathbf{d}(s)) | s \in S^\varepsilon(\mathbf{K})\}. \quad (5.10)$$

Let us now prove the first assertion in those cases. Given $x \in S^\varepsilon(\mathbf{K})$, let $d = (1 + xx^*)^{-1/2}$ and $c = dx$. As noted in the proof of Theorem 2 (cf. also Lemma 4),

$(c \ d) \in W(\mathbf{K})$, $d^* = d$, and $x = d^{-1}c$. Since ${}^t x = -x$, we have $\bar{d} = (1 + x^*x)^{-1/2}$. If we assume that $dx = x\bar{d}$, then ${}^t c = -c$ and $c \cdot {}^t d = dc$, and hence $(c \ d) \in Z^\varepsilon(\mathbf{K})$. By (5.10) we have $(c \ d) = a(c(s) \ d(s))$ with $a \in U_n(\mathbf{K})$ and $s \in S^\varepsilon(\mathbf{K})$, which proves the desired assertion. The assumed equality $dx = x\bar{d}$ can be proved as follows. Put $f = d^{-1}$ and $g = \bar{d}^{-1}$. Then $f^2x = (1 + xx^*)x = x(1 + x^*x) = xg^2$. Suppose $gv = \lambda v$ with $0 \neq v \in M_1^n(\mathbf{K})$ with $\lambda \in \mathbf{R}$. Then $\lambda > 0$ and $f^2xv = \lambda^2xu$, and hence $fxv = \lambda xv = xgv$. Since $M_1^n(\mathbf{K})$ is spanned by all such v 's, we obtain $fx = xg$, and so $dx = x\bar{d}$. This completes the proof.

Let us note that (5.3) and (5.9) are special cases of a general principle

$$G = K \cdot \exp(\mathfrak{p}). \quad (5.11)$$

Here G is a connected Lie group, K is a compact subgroup of G , and \mathfrak{p} is the subset of the Lie algebra \mathfrak{g} of G defined by

$$\mathfrak{p} = \{X \in \mathfrak{g} | d\theta X = -X\},$$

where θ is an analytic automorphism of G of order 2 such that K is a subgroup of $K_\theta = \{g \in G | \theta g = g\}$ containing the identity component of K_θ . The explicit form of θ in our cases is: $\theta g = \lambda g \lambda^{-1}$, $\lambda = \text{diag}[1_n, -1_m]$ with $m = n$ for (5.9). Relation (5.11) is usually stated only for noncompact groups, but actually it is true in general for the following reason. Since G/K is a complete Riemannian manifold, any point on it can be connected to the origin by a geodesic. Now any geodesic passing through the origin is the image of $\{\exp(tX) | t \in \mathbf{R}\}$, $X \in \mathfrak{p}$, under the natural map $G \rightarrow G/K$ (see [1, pp. 208–9]), which proves (5.11).

In the exceptional cases $C^\varepsilon(\mathbf{K})$ is not connected. So let us denote by $C_0^\varepsilon(\mathbf{K})$ the identity component of $C^\varepsilon(\mathbf{K})$, and put

$$W_0^\varepsilon(\mathbf{K}) = (0 \ 1_n)C_0^\varepsilon(\mathbf{K}), \quad Z_0^\varepsilon(\mathbf{K}) = \{(c(s) \ d(s)) | s \in S^\varepsilon(\mathbf{K})\}. \quad (5.12)$$

By Lemma 6(2) $W_0^\varepsilon(\mathbf{K})$ is one of the connected components of $W^\varepsilon(\mathbf{K})$.

To see the nature of our group $C^\varepsilon(\mathbf{K})$ more clearly, let us use the traditional notation by putting

$$\begin{aligned} O(n) &= U_n(\mathbf{R}), & SO(n) &= \{\alpha \in O(n) | \det(\alpha) = 1\}, \\ U(n) &= U_n(\mathbf{C}), & SU(n) &= \{\alpha \in U(n) | \det(\alpha) = 1\}. \end{aligned}$$

If $(\varepsilon, \mathbf{K}) = ((+, -), \mathbf{C})$, it is easy to see that the map $\alpha \mapsto \zeta \alpha \zeta^{-1}$ with

$$\zeta = 2^{-1/2} \begin{pmatrix} 1_n & i1_n \\ i1_n & 1_n \end{pmatrix}$$

gives an isomorphism of $C^{(+, -)}(\mathbf{C})$ onto $O(2n)$. Therefore $C_0^{(+, -)}(\mathbf{C}) = SU(2n) \cap C^{(+, -)}(\mathbf{C}) = \zeta^{-1}SO(2n)\zeta$.

In the case $(\varepsilon, \mathbf{K}) = ((+, +), \mathbf{R})$, equality (4.4) together with a simple relation

$$\frac{1}{2} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} d & c \\ c & d \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} d+c & 0 \\ 0 & d-c \end{pmatrix}$$

shows that

$$\begin{pmatrix} d & c \\ c & d \end{pmatrix} \mapsto (d+c, d-c)$$

gives an isomorphism of $C^{(+, +)}(\mathbf{R})$ onto $O(n) \times O(n)$, and $\det \begin{pmatrix} d & c \\ c & d \end{pmatrix} = \det[(d+c)(d-c)]$. Therefore

$$C_0^{(+, +)}(\mathbf{R}) = \left\{ \begin{pmatrix} d & c \\ c & d \end{pmatrix} \in C^{(+, +)}(\mathbf{R}) \mid \det(d+c) = \det(d-c) = 1 \right\}.$$

Theorem 8. Suppose that $(\varepsilon, \mathbf{K}) = ((+, +), \mathbf{R})$ or $((+, -), \mathbf{C})$. Then

$$W_0^\varepsilon(\mathbf{K}) = U'_n(\mathbf{K})Z_0^\varepsilon(\mathbf{K}), \quad (5.13)$$

$$C_0^\varepsilon(\mathbf{K}) = \{\text{diag}[\bar{a}, a] | a \in U'_n(\mathbf{K})\}E(S^\varepsilon(\mathbf{K})), \quad (5.14)$$

$$Z_0^\varepsilon(\mathbf{K}) = \{(c \ d) \in Z^\varepsilon(\mathbf{K}) | d \in S'_n\}, \quad (5.15)$$

where $U'_n(\mathbf{K})$ denotes the identity component of $U_n(\mathbf{K})$, and S'_n the subset of $S_n(\mathbf{C})$ consisting of the elements with -1 as an eigenvalue with even multiplicity (including the case of zero multiplicity). Moreover, $Z^\varepsilon(\mathbf{K})$ has two connected components, and the component containing $(0 \ 1_n)$ is $Z_0^\varepsilon(\mathbf{K})$.

Proof: We have $S^\varepsilon(\mathbf{K}) = \{-^t y = y \in M_n^n(\mathbf{K})\}$ in the present cases. From our definition of $C^\varepsilon(\mathbf{K})$ we easily see that its Lie algebra is

$$\left\{ \begin{pmatrix} \bar{x} & -y^* \\ y & x \end{pmatrix} \middle| -x^* = x \in M_n^n(\mathbf{K}), y \in S^\varepsilon(\mathbf{K}) \right\}.$$

Therefore (5.14) follows immediately from (5.11). Then (5.13) follows from (5.12) and (5.14). To prove (5.15), let $w \in S^\varepsilon(\mathbf{K})$. Take $q \in U_n(\mathbf{K})$ so that $qww^*q^* = \text{diag}[a_1^2, \dots, a_n^2]$ with $0 \leq a_\nu \in \mathbf{R}$. Then $q\mathbf{d}(w)q^* = \text{diag}[\cos a_1, \dots, \cos a_n]$. By Lemma 7 below we easily see that $\mathbf{d}(w) \in S'_n$, and therefore the left-hand side of (5.15) is contained in the right-hand side. Conversely, suppose $(c \ d) \in Z^\varepsilon(\mathbf{K})$ and $d \in S'_n$. Take $p \in U_n(\mathbf{K})$ so that $pdp^{-1} = \text{diag}[a, -1_s, 1_t]$ as in the proof of Theorem 6. Then s is even. Put

$$h = p^{-1} \text{diag} \left[0_r, \pi \begin{pmatrix} 0 & -1_{s/2} \\ 1_{s/2} & 0 \end{pmatrix}, 0_t' \right] \bar{p}$$

and $\omega = \text{diag}[\bar{p}, p]$. Then $\omega E(h)\omega^{-1} = \text{diag}[\delta, \delta]$ with $\delta = \text{diag}[1_r, -1_s, 1_t]$. Take $y = p^{-1} \text{diag}[\sigma b^{-1}v, 0]\bar{p}$ as in the last part of the proof of Theorem 7. Since $\begin{pmatrix} 0 & -h^* \\ h & 0 \end{pmatrix}$ commutes with $\begin{pmatrix} 0 & -y^* \\ y & 0 \end{pmatrix}$, we have $E(h+y) = E(h)E(y)$. Now $E(h) = \text{diag}[\bar{u}, u]$ with $u = p^{-1}\delta p$, and hence $(\mathbf{c}(h+y), \mathbf{d}(h+y)) = (u\mathbf{c}(y), u\mathbf{d}(y)) = (c, d)$. This completes the proof of (5.15). Now $Z_0^\varepsilon(\mathbf{K})$, being a continuous image of a connected set $S^\varepsilon(\mathbf{K})$, is connected. Suppose $(c \ d) \in Z^\varepsilon(\mathbf{K})$ and $d \notin S'_n$. Take $x \in U'_n(\mathbf{K})$ so that $x dx^{-1} = \text{diag}[-1_s, w]$ with a real diagonal matrix w whose eigenvalues are different from -1 . Then s is odd and $xc \cdot^t x = \text{diag}[0_s^s, z]$ with $w \in M_{n-s}^{n-s}(\mathbf{K})$ for the same reason as in the proof of Theorems 6 and 7. This shows that $(c \ d)$ belongs to the set

$$\{(x^{-1} \text{diag}[0, e]\bar{x}, x^{-1} \text{diag}[-1, f]x) | x \in U'_n(\mathbf{K}), (e \ f) \in Z_{n-1}\}, \quad (5.16)$$

where Z_{n-1} is the set $Z_0^\varepsilon(\mathbf{K})$ defined with $n-1$ in place of n . Since both $U'_n(\mathbf{K})$ and Z_{n-1} are connected, (5.16) is connected. Thus $Z^\varepsilon(\mathbf{K})$ has at most two connected components. To show that it is not connected, take any $(c \ d) \in Z^\varepsilon(\mathbf{K})$. In the proof of Theorem 7, we obtained an expression $pdp^{-1} = \text{diag}[a_1 1_{r_1}, \dots, a_k 1_{r_k}, -1_s, 1_t]$, and also $v_\nu = -\varepsilon_1 v_\nu^\rho \in M_{r_\nu}^{r_\nu}(\mathbf{K})$. Since v_ν is invertible, r_ν must be even. From this we can conclude that the parity of the number of negative eigenvalues of d is a continuous function on $Z^\varepsilon(\mathbf{K})$. This proves the last assertion of our theorem, since both $(0, 1_n)$ and $(0, \text{diag}[-1, 1_{n-1}])$ belong to $Z^\varepsilon(\mathbf{K})$.

Lemma 7. If $w = -^t w \in M_n^n(\mathbf{C})$, then the multiplicity of every nonzero eigenvalue of ww^* is even.

Proof: Take $u \in U_n(\mathbf{C})$ so that $d = uww^*u^*$ is diagonal. Put $z = uw \cdot {}^t u = x + iy$ with $x, y \in M_n^{\mathbf{R}}(\mathbf{R})$. Then ${}^t z = -z$, ${}^t x = -x$, ${}^t y = -y$, and $z\bar{z} = -d$. Since d is real, we have $xy = yx$. Thus x and y are commuting normal matrices, and so simultaneously diagonalizable by an element of $U_n(\mathbf{C})$. Therefore we can find an element v of $U_n(\mathbf{C})$ such that $vzv^* = \text{diag}[c_1, \dots, c_n]$ with $c_n \in \mathbf{C}$. Then $vdv^* = \text{diag}[|c_1|^2, \dots, |c_n|^2]$. Since ${}^t z = -z$, we see that $\{-c_1, \dots, -c_n\}$ coincides with $\{c_1, \dots, c_n\}$ as a whole. Therefore we obtain our lemma.

The decomposition for the group $C^{(-,+)}(\mathbf{K})$ in (5.9) can be expressed in a somewhat different way. We first define a ring-injection $\kappa: M_n^{\mathbf{R}}(\mathbf{H}) \rightarrow M_{2n}^{2n}(\mathbf{C})$ by

$$\kappa(u + vj) = \begin{pmatrix} u & -v \\ \bar{v} & \bar{u} \end{pmatrix} \quad \text{for } u, v \in M_n^{\mathbf{R}}(\mathbf{C}),$$

where j is one of the standard quaternion units. Now it can easily be verified that

$$\begin{pmatrix} d & -c \\ c & d \end{pmatrix} \mapsto \begin{cases} ic + d & (\mathbf{K} = \mathbf{R}), \\ (ic + d, -ic + d) & (\mathbf{K} = \mathbf{C}), \\ i\kappa(c) + \kappa(d) & (\mathbf{K} = \mathbf{H}) \end{cases}$$

gives an isomorphism of $C^{(-,+)}(\mathbf{R})$ onto $U(n)$, $C^{(-,+)}(\mathbf{C})$ onto $U(n) \times U(n)$, and $C^{(-,+)}(\mathbf{H})$ onto $U(2n)$. Then (5.9) can be given in the forms

$$\begin{aligned} U(n) &= O(n)\exp(i \cdot S_n(\mathbf{R})), \\ U(n) \times U(n) &= \{(u, u) | u \in U(n)\} \{(u, u^{-1}) | u \in U(n)\}, \\ U(2n) &= \kappa(U_n(\mathbf{H}))\exp(i \cdot \kappa(S_n(\mathbf{H}))). \end{aligned}$$

6. THINGS THAT ARE HYPERBOLIC. As a natural variation on our theme, we can take the group

$$H_{n,m}(\mathbf{K}) = \{X \in GL_{n+m}(\mathbf{K}) | XI_{n,m}X^* = I_{n,m}\}, \quad I_{n,m} = \text{diag}[1_n, -1_m],$$

in place of $U_{n+m}(\mathbf{K})$, and can still prove corresponding theorems. In this case the matrices whose fractional expression is the question must be restricted to the ball

$$B_n^m(\mathbf{K}) = \{y \in M_n^m(\mathbf{K}) | yy^* < 1_m\}.$$

We put also

$$\begin{aligned} X(\mathbf{K}) &= X_{m,n}(\mathbf{K}) = \{(f \ g) \in M_n^m(\mathbf{K}) \times M_m^m(\mathbf{K}) | gg^* - ff^* = 1_m\}, \\ X'(\mathbf{K}) &= \{(f \ g) \in X(\mathbf{K}) | g^* = g > 0\}, \\ X^\varepsilon(\mathbf{K}) &= X_{n,n}(\mathbf{K}) \cap V^\varepsilon(\mathbf{K}), \quad H^\varepsilon(\mathbf{K}) = G^\varepsilon(\mathbf{K}) \cap H_{n,n}(\mathbf{K}), \\ Y^\varepsilon(\mathbf{K}) &= \{(f \ g) \in X^\varepsilon(\mathbf{K}) | f^\rho = -\varepsilon_1 f, g^* = g > 0\}, \\ \Omega_{n,m}(\mathbf{K}) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_{m+n}(\mathbf{K}) \middle| cc^* < dd^* \right\}, \end{aligned}$$

where $c \in M_n^m(\mathbf{K})$. As an analogue of (4.4) we have

$$H^\varepsilon(\mathbf{K}) = \left\{ \begin{pmatrix} d' & -\varepsilon_1 c' \\ c & d \end{pmatrix} \middle| (c \ d) \in X^\varepsilon(\mathbf{K}) \right\},$$

where $(d' \ c')$ is the same as in (4.4).

Theorem 9. (1) $\Omega_{n,m}(\mathbf{K}) = P_{n,m}(\mathbf{K})H_{n,m}(\mathbf{K})$.

(2) $\Omega_{n,m}(\mathbf{K}) \cap G^\varepsilon(\mathbf{K}) = P^\varepsilon(\mathbf{K})H^\varepsilon(\mathbf{K})$.

(3) $(f \ g) \mapsto g^{-1}f$ gives a one-to-one map of $U_m(K) \setminus X(\mathbf{K})$ onto $B_n^m(\mathbf{K})$.

These assertions correspond to Theorem 3(2), Theorem 5(2), and Theorem 2, and can be proved in a straightforward way.

Let us now define $\mathbf{f}(x)$, $\mathbf{g}(x)$, and $F(x)$ for $x \in M_n^m(\mathbf{K})$ by

$$\mathbf{f}(x) = \sum_{\nu=0}^{\infty} (xx^*)^{\nu} x / (2\nu + 1)!, \quad \mathbf{g}(x) = \sum_{\nu=0}^{\infty} (xx^*)^{\nu} / (2\nu)!,$$

$$F(x) = \exp \left(\begin{pmatrix} 0_n^n & x^* \\ x & 0_m^m \end{pmatrix} \right) \quad (x \in M_n^m(\mathbf{K})).$$

Then we have

$$\mathbf{f}(x)^* = \mathbf{f}(x^*), \quad \mathbf{g}(x)^* = \mathbf{g}(x), \quad \text{and} \quad F(x) = \begin{pmatrix} \mathbf{g}(x^*) & \mathbf{f}(x^*) \\ \mathbf{f}(x) & \mathbf{g}(x) \end{pmatrix}.$$

Theorem 10. *The map $x \mapsto \mathbf{g}(x)^{-1}\mathbf{f}(x)$ gives a one-to-one map of $M_n^m(\mathbf{K})$ onto $B_n^m(\mathbf{K})$. Moreover we have*

$$X(\mathbf{K}) = U_m(\mathbf{K})X'(\mathbf{K}), \quad X'(\mathbf{K}) = \{(\mathbf{f}(x) \quad \mathbf{g}(x)) | x \in M_n^m(\mathbf{K})\},$$

$$H_{n,m}(\mathbf{K}) = (U_n(\mathbf{K}) \times U_m(\mathbf{K}))F(M_n^m(\mathbf{K})).$$

Theorem 11. *The map $s \mapsto \mathbf{g}(s)^{-1}\mathbf{f}(s)$ gives a one-to-one map of $S^e(\mathbf{K})$ onto $B_n^n(\mathbf{K}) \cap S^e(\mathbf{K})$. Moreover we have*

$$X^e(\mathbf{K}) = U_n(\mathbf{K})Y^e(\mathbf{K}), \quad Y^e(\mathbf{K}) = \{(\mathbf{f}(s) \quad \mathbf{g}(s)) | s \in S^e(\mathbf{K})\},$$

$$F(S^e(\mathbf{K})) = \{\alpha \in H^e(\mathbf{K}) | \alpha^* = \alpha\},$$

$$H^e(\mathbf{K}) = \{\text{diag}[u', u] | u \in U_n(\mathbf{K})\}F(S^e(\mathbf{K})),$$

where $u' = \bar{u}$ if $(\varepsilon, \mathbf{K}) = ((\pm, -), \mathbf{C})$ and $u' = u$ otherwise.

All these, except the injectivity of the maps in both theorems, can be proved by the same technique as for the corresponding facts in Section 5. In fact the present case is much easier. To see the injectivity, let $y = g^{-1}f = d^{-1}c$ with $(f \quad g)$ and $(c \quad d)$ in $X'(\mathbf{K})$. Then $d = hg$ with $h \in U_m(\mathbf{K})$, $d^2 = g^2$, and hence $d = g$ and $h = 1$. Suppose $f = \mathbf{f}(x)$ and $g = \mathbf{g}(x)$ with $x \in M_n^m(\mathbf{K})$. Put $T = \begin{pmatrix} 0_n^n & x^* \\ x & 0_m^m \end{pmatrix}$. Then $T^* = T$ and

$$F(x) = \exp(T) = \begin{pmatrix} k & f^* \\ f & g \end{pmatrix} \quad \text{with } k = (1_n + f^*f)^{1/2}.$$

Now it is well known (and easy to prove) that $z \mapsto \exp(z)$ gives a bijection of $S_N(\mathbf{K})$ onto $\{y \in S_N(\mathbf{K}) | y > 0\}$. Therefore x is unique for $(f \quad g)$, which proves the desired injectivity.

REFERENCES

1. S. Helgason, *Differential geometry, Lie groups, and symmetric spaces*, Academic Press, 1978.
2. G. Shimura, Confluent hypergeometric functions on tube domains, *Math. Ann.* 260 (1982), 269–302.
3. C. L. Siegel, Einführung in die Theorie der Modulformen n -ten Grades, *Math. Ann.* 116 (1939), 617–657 (= Gesammelte Abhandlungen II, No. 32).
4. C. L. Siegel, On the theory of indefinite quadratic forms, *Ann. of Math.* 45 (1944), 576–622 (= Gesammelte Abhandlungen II, No. 45).

Department of Mathematics
Princeton University
Princeton, New Jersey 08544-1000

Noether Lasker Primary Decomposition Revisited

Barbara L. Osofsky

There are many theorems which are absolutely basic for the study of particular areas of mathematics. Some of these theorems work their way into the standard undergraduate curriculum, but many are reserved for specialized graduate work. In this note, we present a theorem usually reserved for a first graduate course in commutative algebra in a way that undergraduates studying abstract algebra might find congenial. The theorem is the Noether-Lasker Primary Decomposition Theorem, which may be thought of as a way to extend a version of prime factorization from the integers to a much larger class of rings, including polynomials in several variables over a field. Traditionally, the approach taken to this theorem has been arithmetical, i.e. using the multiplicative properties of the ring. Although we refer to the arithmetic of the integers, our approach is more in the spirit of vector spaces. There are a large number of exercises to help the reader get involved in understanding the material. It is not cheating to get help on these exercises.

We assume that the reader has had a first abstract algebra course, including the basic homomorphism theorems for groups and rings, some study of prime factorization for a Euclidean domain such as the integers or polynomials over a field, and at least the definition of a prime ideal in a commutative ring. The other necessary concepts from commutative algebra will be given here. The reader should also have seen the definition of vector space over a field and worked with bases and linear independence in a vector space. That is the background we build on. Although we prove a generalized version of the primary decomposition for finitely generated modules over a commutative Noetherian ring, including uniqueness of primary components if and only if the prime is a minimal associated prime, the reader does not have to know what those words mean to understand our main theorem. The major idea in the proof is contained in a useful module-theoretic definition. After we prove the theorem, we define all the concepts in the Noether-Lasker Theorem so we can specialize our theorem to the case of interest in algebraic geometry and commutative algebra.

A (left) module over a ring R with identity 1 is a system $\langle M, +, \cdot \rangle$ that satisfies the axioms for a vector space with elements of R as scalars written on the left. Any Abelian group is a module over the ring \mathbb{Z} of integers, where scalar multiplication of x by a positive integer n means adding x to itself n times. Ring multiplication induces a module structure on the ring itself. Indeed, if I is any left ideal of the ring R , then the quotient group R/I is a module over R under the induced scalar multiplication $s \cdot (r + I) = sr + I$.

A module homomorphism $\varphi: M \rightarrow N$ between two R -modules satisfies the axioms for a linear transformation. To make the notation look more like the associative law rather than the commutative law, we will write our functions to the right of their arguments, on the side opposite the scalars. Although it does not

matter in this paper, this convention simplifies things elsewhere when non-commutative rings are being studied. For example, when writing down the matrix of a linear transformation T with the matrix on the left and column vectors on the right, you write down the equations

$$T(e_i) = \sum_j a_{j,i} e'_j$$

and set the matrix of T equal to the matrix $[a_{i,j}]$. You are forced to transpose the matrix to have the product (matrix of T) \times (matrix of T') correspond to the composition of T and T' . Writing functions on the right, so $T \circ T'$ means first T and then T' , the equations would read

$$(e_i)T = \sum_j a_{i,j} e'_j$$

and you would not need the transpose. For general modules, there are no matrices associated with homomorphisms, but there are definite benefits to avoiding the equivalent of that transpose.

With this convention on sides, the defining property of a module homomorphism $\varphi: M \rightarrow N$ is that $(r \cdot x + s \cdot y)\varphi = r \cdot (x)\varphi + s \cdot (y)\varphi$ for all $x, y \in M$ and $r, s \in R$. The homomorphism φ is an isomorphism if it is one-to-one and onto. Just as in other algebraic structures such as groups, rings, vector spaces, an isomorphism is essentially a renaming of the elements of the module M .

Definition. A module M is called *cyclic* provided $M = \{r \cdot x | r \in R\}$ for some $x \in M$.

Exercise 1. Any cyclic module $R \cdot x$ is isomorphic to the quotient module R/I , where $I = \{r \in R | r \cdot x = 0\}$.

With that background material listed, we proceed to define the main concept we will need in this note.

Definitions. A non-zero module M is *uniform* if any two non-zero submodules have non-zero intersection. M *contains enough uniforms* provided every non-zero submodule contains a uniform submodule. If U and V are uniform, we say U is *subisomorphic* to V and write $U \sim V$ provided U and V contain non-zero isomorphic submodules. M is called *primary* if M has enough uniforms and any two uniform submodules of M are subisomorphic.

To illustrate the concept of uniform module, we show that the module \mathbb{Z} is a uniform module over the integers \mathbb{Z} . Let I and J be two non-zero ideals of \mathbb{Z} , and pick a non-zero $m \in I$ and a non-zero $n \in J$. Then the non-zero element $mn \in I \cap J$, so $I \cap J \neq 0$. Note that \mathbb{Z} or any uniform module must be primary.

Exercise 2. Prove \sim is an equivalence relation on the class of all uniform modules.

Notation. We let $[U]$ denote the equivalence class of the uniform module U .

Exercise 3. Any commutative domain is uniform as a module over itself.

Exercise 4. A vector space over a field is uniform if and only if it is one-dimensional. Any non-zero vector space contains enough uniforms and is primary.

Exercise 5. A cyclic \mathbb{Z} -module, $\mathbb{Z}/n\mathbb{Z}$, is uniform if and only if n is 0 or a power p^i of a prime p .

Exercise 6. Any \mathbb{Z} -module, i.e. any Abelian group, contains enough uniforms.

Exercise 7. Every module over a ring R contains enough uniforms if and only if every non-zero cyclic module R/I contains a uniform submodule.

There is an old proverb that a picture is worth a thousand words. In mathematics notation can also be worth many words. So we introduce some very special notation to save some words.

Notation. Let the non-zero module M contain enough uniforms. For each class $[U]$ of uniforms, set

$$\mathcal{F}_{[U], M} = \{K \subseteq M \mid K \text{ a submodule which contains no submodule in } [U]\}.$$

Thus writing $K \in \mathcal{F}_{[U], M}$ is a short way of saying that K is a submodule of M which contains no uniform submodule subisomorphic to U . $\mathcal{F}_{[U], M}$ is never empty as the zero submodule is always in it.

Exercise 8. Let M contain enough uniforms. Then for any uniform module U and any $L \in \mathcal{F}_{[U], M}$, there is a maximal element of $\mathcal{F}_{[U], M}$ which contains L (and which possibly might be 0 or M).

Hint. This requires the set theoretic axiom Zorn's lemma.

We now have enough preliminaries to state and prove our Theorem A. Exercise 8 shows that the module $N_{[U]}$ in Theorem A must exist. Part (b) of Theorem A justifies calling the set of all $N_{[U]}$ which are distinct from M the primary components of 0, and their intersection a primary decomposition of 0 in M .

Theorem A. Let the non-zero module M contain enough uniforms. For each class $[U]$ of uniforms, let $N_{[U]}$ be any maximal element of $\mathcal{F}_{[U], M}$. Then

- (a) $\bigcap_{[U]} N_{[U]} = 0$.
- (b) If $0 \neq M/N_{[U]}$, then $M/N_{[U]}$ is primary, that is, it contains enough uniforms and any uniform submodule of $M/N_{[U]}$ is in $[U]$.
- (c) $\mathcal{F}_{[U], M}$ has more than one maximal element \Leftrightarrow there is a $K \in \mathcal{F}_{[U], M}$ and a non-zero homomorphism φ from K to $V \subseteq M$ where V is a uniform module in $[U]$.

Proof: To see (a), let $0 \neq x \in M$. Then $R \cdot x$ contains a uniform module V . By definition, $N_{[V]}$ cannot have a non-zero submodule isomorphic to a submodule of V . Thus $V \cap N_{[V]} = 0$ and $x \notin N_{[V]}$.

To see (b), we draw a picture of the situation in Figure 1 and then give the argument.

$$\begin{array}{ccc}
 & V & \\
 & | \cap & \\
 & W \subseteq M & \\
 \downarrow \nu & & \downarrow \nu \\
 0 \neq K \subseteq M/N_{[U]} & &
 \end{array}$$

Figure 1

Let K be any non-zero submodule of $M/N_{[U]}$ and W the largest submodule of M mapping onto K modulo $N_{[U]}$. Since W properly contains $N_{[U]}$, by maximality of $N_{[U]}$ in $\mathcal{F}_{[U], M}$, W must contain a uniform submodule V subisomorphic to U . Look at the module $V \cap N_{[U]}$. If it is not zero, then it is a uniform module contained in $N_{[U]}$ and it is subisomorphic to V which is in $[U]$. By definition of $N_{[U]}$, no such non-zero module exists, so $V \cap N_{[U]} = 0$. Restrict the quotient map $\nu: M \rightarrow M/N_{[U]}$ to V (i.e. look at the left column of Figure 1). This restriction has kernel $V \cap N_{[U]} = 0$ and so is one-to-one. Then K contains a submodule $(V)\nu$ isomorphic to $V \in [U]$.

For \Rightarrow in (c), assume that $K_{[U]}$ and $L_{[U]}$ are two distinct maximal elements of $\mathcal{F}_{[U], M}$. We again draw a picture to aid in following the argument.

$$\begin{array}{ccc}
 H \subseteq L_{[U]} \subseteq M & & \\
 \downarrow \nu & \downarrow \nu & \nearrow \\
 V \subseteq M/K_{[U]} & &
 \end{array}$$

Figure 2

Since $L_{[U]}$ is not contained in $K_{[U]}$, the natural map ν from M to $M/K_{[U]}$ induces a non-zero map on $L_{[U]}$. By (b), the image of ν restricted to $L_{[U]}$ contains enough uniforms, and all of its uniform submodules are in $[U]$. Restrict ν even further to the preimage H in $L_{[U]}$ of a uniform submodule $V \subseteq (L_{[U]})\nu$ (i.e. look at the left hand column of Figure 2). Then ν restricted to H is the non-zero map required in the \Rightarrow part of (c).

For the \Leftarrow direction of (c) assume there is a non-zero map φ from a submodule $K \in \mathcal{F}_{[U], M}$ to a submodule $V \subseteq M$, $V \in [U]$. Pick $x \in K$ with $0 \neq (x)\varphi$. For the moment we assume that $r \cdot x = 0 \Leftrightarrow r \cdot (x - (x)\varphi) = 0$.

By a basic homomorphism theorem and our assumption, both $R \cdot x$ and $R \cdot (x - (x)\varphi)$ are isomorphic to the same quotient module $R/\{r \in R \mid r \cdot x = 0\}$. Since $K_1 = R \cdot x$ is in $\mathcal{F}_{[U], M}$, so is the isomorphic module $K_2 = R \cdot (x - (x)\varphi)$. However, the sum $K_1 + K_2$ contains the module $R \cdot (x)\varphi \in [U]$, so no submodule $L \in \mathcal{F}_{[U], M}$ contains both K_1 and K_2 . By Exercise 8, each K_i is contained in a maximal element of $\mathcal{F}_{[U], M}$ so there must be more than one maximal element in $\mathcal{F}_{[U], M}$.

Now we prove our assumption that $r \cdot x = 0 \Leftrightarrow r \cdot (x - (x)\varphi) = 0$. Since $R \cdot x \in \mathcal{F}_{[U], M}$ and $R \cdot (x)\varphi \in [U]$, $R \cdot x \cap R \cdot (x)\varphi = 0$. Then $0 = r \cdot (x - (x)\varphi)$ implies $0 = r \cdot x - r \cdot (x)\varphi$ by the distributive law for scalar multiplication. Bringing $r \cdot (x)\varphi$ over to the other side shows that $r \cdot (x)\varphi = r \cdot x \in R \cdot x \cap R \cdot (x)\varphi = 0$, so any r with $r \cdot (x - (x)\varphi) = 0$ has $r \cdot x = 0$. Conversely, if $0 = r \cdot x$, then $0 = (r \cdot x)\varphi = r \cdot ((x)\varphi)$ so $0 = r \cdot x - r \cdot (x)\varphi = r \cdot (x - (x)\varphi)$. \square

If $R = F$ is a field, there is only one subisomorphism class of uniform modules, namely the class $[F]$ of the field as a module over itself. In this case $N_{[F]} = 0$ for any non-zero vector space M , and the theorem does not yield much information.

What does this theorem say about modules over \mathbb{Z} ? Exercise 5 asked for a proof that the only cyclic uniform modules over \mathbb{Z} are isomorphic either to \mathbb{Z} or to $\mathbb{Z}/p^i\mathbb{Z}$ for some prime p . Exercise 9 asks you to specialize the Theorem to an Abelian group M . Let $\tau(M)$ denote the torsion subgroup of M , that is, the set of all elements in M of finite order.

Exercise 9. *Let M be an Abelian group. Show*

- (a) *For each prime p there exists a submodule $N_p \subseteq M$ such that every element of M/N_p is of order a power of p . Furthermore,*

$$\left(\bigcap_{p \text{ prime}} N_p \right) \cap \tau(M) = 0. \quad (*)$$

- (b) *If $\tau(M) \neq M$ then each N_p contains a subgroup isomorphic to \mathbb{Z} . If in addition $M/N_p \neq 0$, then there is a non-zero map from this subgroup to a uniform submodule of M isomorphic to $\mathbb{Z}/\mathbb{Z} \cdot p$.*
(c) *If $\tau(M) \neq M$, then for any p with $N_p \neq M$, N_p is not a unique submodule of M .*
(d) *If $\tau(M) = M$, then all the N_p are unique.*

Exercise 9 forms a bridge to the theorem mentioned in the first paragraph. We next define some of the terms used there.

Definition. A module M is called *Noetherian* provided every submodule of M is finitely generated, that is, if N is a submodule of M , then there exists a finite set $\{a_1, \dots, a_n\} \subseteq N$ such that $N = \{\sum_{j=1}^n r_j \cdot a_j \mid r_j \in R\}$. A ring is Noetherian if it is Noetherian as a module over itself. M has the *maximum condition* if for any non-empty family \mathcal{F} of submodules of M there is a maximal element of \mathcal{F} .

Exercise 10. \mathbb{Z} is Noetherian and has the maximum condition.

Fact. The rings that arise in the study of algebraic geometry and algebraic number theory are commutative Noetherian rings. It is a well known theorem attributed to Hilbert that polynomials over a Noetherian ring are again Noetherian. Thus the ring $F[X_1, \dots, X_n]$ of polynomials in n variables over a field F is Noetherian. So is $\mathbb{Z}[X_1, \dots, X_n]$.

Exercise 11. M has the maximum condition $\Leftrightarrow M$ is Noetherian.

Hint. A third condition, called the ascending chain condition (acc), is usually given as another condition equivalent to Noetherian. It states that for any ascending chain

$$K_0 \subseteq K_1 \subseteq \dots \subseteq K_n \subseteq \dots$$

of submodules of M there must be an l with $K_i = K_l$ for all $i \geq l$. It might be useful in doing Exercise 11 to show maximum condition \Rightarrow Noetherian \Rightarrow acc \Rightarrow maximum condition.

Exercise 12. If a ring R is Noetherian, then every submodule of a finite direct sum of copies of the R -module R is also finitely generated.

Hint. Use finite induction on the number of summands, the projection onto the last summand, and the one-to-one correspondence homomorphism theorem.

Exercise 13. If M is a finitely generated module over a Noetherian ring, and \mathcal{F} is a non-empty family of submodules of M , then there is a maximal element of \mathcal{F} .

What does all this have to do with uniform modules?

Lemma 1. Let M be a Noetherian module or any module over a Noetherian ring. Then M has enough uniforms.

Proof: Let N be a non-zero submodule of M , $0 \neq x \in N$. By Exercise 11, the maximum condition is equivalent to Noetherian. We observe that any submodule of a module with maximum condition must have maximum condition and any quotient of a Noetherian module is Noetherian. Thus either hypothesis on M implies $R \cdot x$ has maximum condition. To use this fact, we must find the correct non-empty family to have a maximal element and then use that maximality.

Let

$$\mathcal{F} = \{K \text{ a submodule of } R \cdot x \mid \text{there is a submodule } L \neq 0 \text{ of } R \cdot x \text{ with } L \cap K = 0\}.$$

\mathcal{F} is not empty since $\{0\} \in \mathcal{F}$. Let K_0 be a maximal element of \mathcal{F} , and let U be a non-zero submodule of $R \cdot x$ with $U \cap K_0 = 0$. Assume U is not uniform. Then it contains non-zero submodules H and J such that $H \cap J = 0$. By Exercise 14 below, $H \cap (J + K_0) = 0$ so $J + K_0 \in \mathcal{F}$. Since $J \not\subseteq K_0$, $J + K_0$ properly contains K_0 , contradicting the maximality of K_0 . We conclude that U must be uniform. \square

Exercise 14. Complete the details in the proof of Lemma 1 by showing that if $H \cap J = 0$ and $(H + J) \cap K_0 = 0$ then $H \cap (J + K_0) = 0$.

Hint. You must use both of the hypotheses. Start with $h = j + k \in H \cap (J + K_0)$ and manipulate to show $h = j = k = 0$.

We now proceed to do some traditional commutative algebra. We first state the Noether-Lasker Primary Decomposition Theorem, which is what motivated this note.

Theorem B [Noether-Lasker]. Let M be a finitely generated module over a commutative Noetherian ring R . Then there exists a finite set $\{N_i \mid 1 \leq i \leq l\}$ of submodules of M such that:

- (a) $\bigcap_{i=1}^l N_i = 0$ and $\bigcap_{i \neq i_0} N_i$ is not contained in N_{i_0} for all $1 \leq i_0 \leq l$.
- (b) Each quotient M/N_i is primary for some prime P_i .
- (c) The P_i are all distinct for $1 \leq i \leq l$.
- (d) The primary component N_i is unique $\Leftrightarrow P_i$ does not contain P_j for any $j \neq i$.

This theorem looks somewhat different than Theorem A. There are several “problems” with it.

- The Noether-Lasker Theorem talks about prime ideals, not uniform modules.
- The uniqueness portion of Theorem A seems to have nothing to do with prime ideals being contained in other prime ideals.
- In the standard Noether-Lasker Theorem, uniqueness in (d) is asserted for any N_i satisfying (a), (b) and (c), not only maximal elements of $\mathcal{F}_{[U], M}$.
- Theorem A says nothing about finiteness.
- The usual definition of the word primary is not the same as our definition. It is actually a special case of our definition.

In the rest of this paper we will address these “problems”.

Lemma 2. *Let M have enough uniforms and let $\{N_i | 1 \leq i \leq k < \infty\}$ satisfy*

- (a) $\bigcap_{i=1}^k N_i = 0$.
- (b) M/N_i is primary with uniforms all in $[U_i]$.
- (c) $[U_i] \neq [U_j]$ if $i \neq j$.

Then each $N_i \in \mathcal{F}_{[U_i], M}$. Moreover, if $\mathcal{F}_{[U_i], M}$ has only one maximal element which is distinct from M , then N_i is that maximal element.

Proof: Let U be a uniform submodule of M with $U \in [U_i]$. Then for $j \neq i$, U cannot embed in M/N_j . Thus $U \cap N_j \neq 0$ for $j \neq i$. Since U is uniform, by finite induction $K_i = \bigcap_{j \neq i} (U \cap N_j) \neq 0$. But $K_i \cap (U \cap N_i)$ is equal to 0, so by the uniformity of U , $U \cap N_i = 0$. Thus $N_i \in \mathcal{F}_{[U_i], M}$.

If $\mathcal{F}_{[U_i], M}$ contains a unique maximum element $L_{[U_i]}$ which is not M , then M contains a uniform $U \in [U_i]$ and $L_{[U_i]}/N_i$, which is $[U_i]$ -primary, cannot have a non-zero map from a submodule to U . Thus N_i must equal $L_{[U_i]}$. \square

Note that in Lemma 2, if N_i does not contain $\bigcap_{j \neq i} N_j$, then $M \notin \mathcal{F}_{[U_i], M}$. Can you show that Lemma 2 need not be true if the number of N_i is not finite?

Hint. Look at the Abelian group $\mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Q}/\mathbb{Z}$.

What do uniform modules over a commutative Noetherian ring R look like? If P is a prime ideal of R , then R/P is a domain and hence a uniform R/P -module. Since multiplication by an element $r \in R$ gives the same result as multiplication by $r + P \in R/P$, R/P is also a uniform R -module. We next show that every uniform R -module is in the equivalence class $[R/P]$ for precisely one prime ideal P .

Notation. Let M be any R -module. For $x \in M$ we let $(0 : x)$ denote the annihilator of x in R , that is,

$$(0 : x) = \{r \in R | r \cdot x = 0\}.$$

Lemma 3. *If R is commutative, then any ideal P maximal in $\{(0 : x) | 0 \neq x \in M\}$ is prime.*

Proof: We first note that P cannot equal R since $1 \notin P$. Let $0 \neq x_0 \in M$ have $P = (0 : x_0)$. If $ab \in P$ and $b \notin P$, then $0 \neq b \cdot x_0$, and $(0 : b \cdot x_0) \supseteq R \cdot a + P$. By maximality of P , $P = R \cdot a + P$, so $a \in P$. Thus P is prime. \square

Lemma 4. *Let U be a uniform module over a commutative Noetherian ring R . Then U contains a submodule isomorphic to R/P for precisely one prime ideal P .*

Proof: By Exercise 11, $\{(0 : x) \mid 0 \neq x \in U\}$ has a maximal element. By Lemma 3, a maximal element in this set of annihilators of elements of U is prime. If $P = (0 : x_0)$ is such a maximal annihilator, then U contains the submodule $R \cdot x_0$ which is isomorphic to R/P .

We observe that every non-zero cyclic submodule of R/P , for R commutative and P a prime ideal of R , is isomorphic to R/P . If I is any ideal distinct from P then R/I is not isomorphic to R/P . Hence if U and V are uniform R -modules containing submodules isomorphic to R/P and R/Q respectively where P and Q are distinct primes, then $[U] \neq [V]$. Thus any uniform module contains a submodule isomorphic to R/P for one and only one prime P . \square

We next look at the uniqueness portion of the Noether-Lasker Theorem.

Definition. If R is a commutative Noetherian ring and P is a prime ideal of R , then P is *associated with* the module M provided $P = (0 : x)$ for some $x \in M$. We will call MP -*primary* if the prime ideal P is associated with M and no other prime is.

Exercise 15. *If R is a commutative Noetherian ring and P is a prime ideal of R , then an R -module M is P -primary in the above sense if and only if M has enough uniforms and every uniform submodule of M is in $[R/P]$.*

Lemma 5. *Let R be commutative Noetherian, and let M have the set of associated prime ideals $\{P_i \mid i \in \mathcal{J}\}$. Let $x \in M$ and $p \in \bigcap_{i \in \mathcal{J}} P_i$. Then there is a nonnegative integer n such that $p^n \cdot x = 0$.*

Proof: Let $I = (0 : p^n \cdot x)$ be a maximal element of $\{(0 : p^i \cdot x) \mid i \in \mathbb{N}\}$. If $p^n \cdot x \neq 0$, by Lemma 4 some submodule of $Rp^n \cdot x$, say $Rsp^n \cdot x$, is isomorphic to R/P_i for some $i \in \mathcal{J}$. Then $p^n s \cdot x = sp^n \cdot x \neq 0$, but since $p \in P_i$ we have $p \cdot (p^n s \cdot x) = pp^n s \cdot x = 0$. Then $(0 : p^{n+1} \cdot x) \supseteq I + Rs$ which properly contains I , contradicting the maximality of I . Hence our assumption that $p^n \cdot x \neq 0$ must be wrong, i.e. $p^n \cdot x = 0$. \square

Corollary 6. *For R commutative Noetherian, let the set of primes $\{P_1, \dots, P_k\}$ associated with M be finite, and let Q be any prime which does not contain any of the P_i . Then the only homomorphism from M to R/Q is the zero homomorphism.*

Proof: For each i with $1 \leq i \leq k$ let $p_i \in P_i$, $p_i \notin Q$. Then $s = \prod_{i=1}^k p_i \notin Q$. Let φ be a homomorphism from M to R/Q , $x \in M$. By Lemma 5 there is an n such that $s^n \cdot x = 0$. Then $0 = (0)\varphi = (s^n \cdot x)\varphi = s^n \cdot ((x)\varphi)$. But Q is a prime ideal and $s^n \notin Q$, so $s^n \cdot ((x)\varphi) = (0 + Q)$ in R/Q implies $((x)\varphi) = (0 + Q)$ in R/Q . \square

The next property we must examine to show that our theorem reduces to the standard primary decomposition of a finitely generated module over a commutative Noetherian ring is the finiteness of the set of associated primes.

We first give a definition which generalizes the concept of linear independence for vector spaces. A set \mathcal{B} of elements in a vector space is called linearly

independent provided the condition $\sum r_i \cdot b_i = 0$, where the b_i are distinct elements of \mathcal{B} , implies that each $r_i = 0$. By convention, the empty set \emptyset is considered independent and indeed a basis for the 0 vector space.

Definition. A family $\mathcal{B} = \{U_i \subseteq M\}$ of submodules of M is called *independent* if, whenever you have $\sum_i u_i = 0$ with the u_i from distinct U_i , then each $u_i = 0$. The empty set of submodules is considered independent.

Exercise 16. Let M be a module. Then there exists a maximal (possibly empty) independent set of uniform submodules of M .

Hint. This uses Zorn's Lemma. This is precisely the way in which one proves that any vector space has a basis even if it is not finite-dimensional.

Lemma 7. Let M be a non-zero finitely generated module over a commutative Noetherian ring R . Then there are only a finite number of primes associated with M .

Proof: By Exercise 16, there is a maximal independent set $\mathcal{B} = \{U_i | i \in \mathcal{I}\}$ of uniform submodules of M . By Lemma 1, which says that M contains enough uniforms, \mathcal{B} is not empty. Set $K = \sum_{i \in \mathcal{I}} U_i$. Since K is finitely generated, some finite subsum of $\sum_i U_i$ must contain all of the generators. Then K must equal that finite subsum, say $K = \sum_{i=1}^l U_i$. By the independence of \mathcal{B} , no other U_j can have a non-zero element in common with $\sum_{i=1}^l U_i$, so $\mathcal{B} = \{U_i | 1 \leq i \leq l\}$ is a finite set.

Each U_i has a single associated prime P_i . Let $Q = (0 : x_0)$ be an associated prime for M . Then $R \cdot x_0$ is a uniform module. Just as we have for vector spaces,

$$\text{either } R \cdot x_0 \cap \sum_{i=1}^l U_i \neq 0 \text{ or } \{R \cdot x_0, U_1, \dots, U_l\} \text{ is independent.}$$

By maximality of \mathcal{B} , $R \cdot x_0 \cap K \neq 0$. Let $0 \neq k = \sum_{i=1}^l u_i \in R \cdot x_0 \cap K$. Then $Q = (0 : k) = \cap_{i=1}^l (0 : u_i) = \cap_{u_j \neq 0} P_j$. This tells us that $Q \subseteq P_j$ for each j with $u_j \neq 0$. If, for each non-zero u_j , there is an element $p_j \in P_j$, $p_j \notin Q$, then $\prod_{u_j \neq 0} p_j$ must belong to Q so one of the p_j must be in Q , a contradiction. We conclude that $Q \supseteq P_j$ for some such j . Thus $Q \in \{P_1, \dots, P_l\}$. \square

We now come to our last “problem”. Even our definition of P -primary is not the traditional definition used in classical commutative algebra texts. For the record, traditionally M is called primary provided that, whenever $x \in M$ and $r \in R$ satisfy $r \cdot x = 0$, either $x = 0$ or there is a non-negative integer n such that $r^n \cdot M = 0$. For finitely generated modules over commutative Noetherian rings, the definitions are equivalent. Otherwise ours is more general. The second condition in Lemma 8 below is this traditional definition of primary. It has a much more arithmetic flavor than our definition.

Lemma 8. Let M be a non-zero finitely generated module over a commutative Noetherian ring R . Then M is a P -primary module for some prime ideal P of R if and only if for all $0 \neq x \in M$ and $r \in R$, if $r \cdot x = 0$, then there exists an $n \in \mathbb{N}$ such that $r^n \cdot y = 0$ for all $y \in M$.

Proof: For the only if direction, assume M is P -primary for some prime P . Let $0 \neq x \in M$ and $r \in R$ have $r \cdot x = 0$. By Lemma 3, there is an $s \in R$ with

$(0:s \cdot x)$ a prime ideal. Since M is P -primary, $(0:s \cdot x)$ must equal P . Since $r \cdot (s \cdot x) = s \cdot r \cdot x = 0$, $r \in P$. By Lemma 5, some power of each element of P annihilates every element of M . Since M is finitely generated, $M = \sum_1^k R \cdot x_j$ for some finite set $\{x_j | 1 \leq j \leq k\}$. If $r^{n_j} \cdot x_j = 0$, we leave it as an exercise that $r^{(n_1 + \dots + n_k)} \cdot M = 0$.

For the if direction, assume for all $0 \neq x \in M$ and $r \in R$, if $r \cdot x = 0$, then there exists an $n \in \mathbb{N}$ such that $r^n \cdot y = 0$ for all $y \in M$. Set

$$P = \{s \in R | \text{there is an } n \in \mathbb{N} \text{ with } s^n \cdot M = 0\}.$$

Clearly $0 \in P$ so $P \neq \emptyset$. Equally clearly, $1 \notin P$. Let $s, \hat{s} \in P$, and $r \in R$. If $s^n \cdot M = 0$ and $\hat{s}^{\hat{n}} = 0$, then $(s + \hat{s})^{(n+\hat{n})} \cdot M = 0$ and $(rs)^n \cdot M = 0$. Thus P is an ideal not equal to R . Let U be any uniform submodule of M . Let Q be an associated prime of M , that is, $Q = (0:x)$ for some non-zero $x \in M$ and Q is prime. By hypothesis, every element of Q is in P , that is, $Q \subseteq P$. If $s \in P$, then some power of s is in $(0:x) = Q$ which is a prime ideal. Thus $s \in Q$, so $P \subseteq Q$. We conclude that $P = Q$ is the unique prime associated to M , so M is P -primary. \square

To enable the reader to apply the ideas in this paper, we include one last exercise.

Exercise 17 [Final Exam]. Let $R = \mathbb{R}[X, Y, Z]$ be the ring of polynomials in 3 indeterminates over the reals. Let I be the ideal of R generated by $\{X^2, XY, XZ^2\}$, that is,

$$I = \{f_1(X, Y, Z)X^2 + f_2(X, Y, Z)XY + f_3(X, Y, Z)XZ^2$$

$$| f_i(X, Y, Z) \in \mathbb{R}[X, Y, Z]\}$$

Find the primes associated to the module $M = R/I$, a set of primary components $N_{[R/P]}$, and at least two different primary components for each prime for which that is possible.

Department of Mathematics
Rutgers University
New Brunswick, NJ 08903
osofsky@math.rutgers.edu

DeMorgan was explaining to an actuary what was the chance that a certain proportion of some group of people would at the end of a given time be alive; and quoted the actuarial formula, involving π , which, in answer to a question, he explained stood for the ratio of the circumference of a circle to its diameter. His acquaintance, who had so far listened to the explanation with interest, interrupted him and exclaimed, "My dear friend, that must be a delusion, what can a circle have to do with the number of people alive at a given time?"

—Walter William Rouose Ball (1850–1925)

Mathematical Recreations and Problems.

London: 1896, p. 180.

Elementary Infinite Sources of Non-Unique Factorization Rings

S. Stein and S. Szabó

When students first meet unique-factorization rings, usually \mathbb{Z} , $\mathbb{Z}[i]$, and $F[x]$, where F is a field, they may get the impression that non-unique-factorization rings are rare. After all, the text usually presents only one or two examples of such rings. Actually, with just a little more effort, we can offer an infinite number of such rings. We describe two ways to do this. The first could be presented at the beginning of the discussion of unique factorization; the second depends on the fact that if R is a unique-factorization ring, then factorizations in $R[x]$ and $F[x]$, where F is the field of fractions of R , are the same up to constant factors.

For our first source of examples let $t \in \mathbb{Z}$ be an odd, negative integer, $t \leq -3$. Then $\mathbb{Z}[\sqrt{t}]$ is a non-unique-factorization ring. To show this, consider the product $(1 + \sqrt{t})(1 - \sqrt{t}) = 1 - t$, which is an even integer greater than or equal to 4.

Let $\phi(a + b\sqrt{t}) = (a + b\sqrt{t})(a - b\sqrt{t}) = a^2 - tb^2$, the usual norm. Using this norm, one shows that 2 has no non-trivial factorization in $\mathbb{Z}[\sqrt{t}]$. However, it is clear that 2 divides neither $1 + \sqrt{t}$ nor $1 - \sqrt{t}$ in $\mathbb{Z}[\sqrt{t}]$, though it divides their product. Thus $1 - t$ has at least two different factorizations as products of primes, so $\mathbb{Z}[\sqrt{t}]$ is a non-unique-factorization ring. This method does not work when t is an odd positive integer, since, for example, $\mathbb{Z}[\sqrt{3}]$ and $\mathbb{Z}[\sqrt{7}]$ are unique-factorization rings. (They are some of the rings of algebraic integers that are Euclidean.)

Our second source of examples depends on the fact that if R is an integral domain such that $R[x]$ contains a polynomial whose factorization into primes in $R[x]$ is essentially different from its factorization in $F[x]$, then R is a non-unique-factorization ring.

Using this observation, we show that if $t \in \mathbb{Z}$ is not a square and is of the form $4k + 1$, $k \in \mathbb{Z}$, then $R = \mathbb{Z}[\sqrt{t}]$ is a non-unique-factorization ring.

Consider the polynomial $Q(x) = x^2 + x + (1 - t)/4$, which lies in $R[x]$. Let F be the field of fractions of R . $Q(x)$ has in $F[x]$ the factorization

$$\left[x + (1/2 + \sqrt{t}/2)\right]\left[x + (1/2 - \sqrt{t}/2)\right].$$

If R were a unique-factorization-ring and $Q(x)$ were reducible in $R[x]$, say $Q(x) = A(x)B(x)$, then there would be integers \underline{a} and b such that

$$A(x) = (a + b\sqrt{t})(x + (1/2 + \sqrt{t}/2))$$

and

$$B(x) = [(a - b\sqrt{t})/(a^2 - tb^2)](x + (1/2 - \sqrt{t}/2)).$$

Thus $a^2 - tb^2$ divides both \underline{a} and b , hence $(a^2 - tb^2)^2$ divides $a^2 - tb^2$, so $a^2 - tb^2 = \pm 1$, implying that one of \underline{a} and b is odd, the other even. The constant term in $A(x)$ is $(a + tb)/2 + (a + b)\sqrt{t}/2$, which shows that $a + b$ is even,

contradicting the fact that a and b have opposite parities. Hence R is a non-unique-factorization ring.

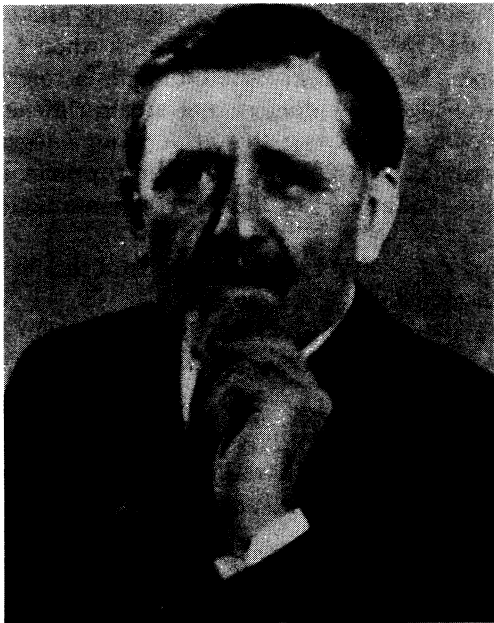
Note that if $e \neq \pm 1$ is an odd integer, then $\mathbb{Z}[\sqrt{e^2d}]$ is properly contained in $\mathbb{Z}[\sqrt{d}]$. Therefore there is an infinite descending chain of non-unique-factorization rings. (These rings are contained in the ring of algebraic integers in $\mathbb{Q}(\sqrt{d})$, which, for $d = 4k + 1$, also contain numbers of the form $a + b\sqrt{d}$, where both a and b are fractions with denominator 2.)

The examples illustrate two general techniques. In the first, a homomorphism is used to project R on a smaller structure, \mathbb{Z} . In the second, R is injected into a larger structure, $R[x]$, by an isomorphism. The second example can also be analyzed with the aid of the “rational root” theorem for R .

Mathematics Department
University of California, Davis
Davis, CA 95616-8633
sstein@math.ucdavis.edu

Mathematics Department
University of Bahrain
P.O. Box 32038
Isa Town, State of Bahrain
esø5ø@isa.cc.uob.bh

PICTURE PUZZLE
(from the collection of Paul Halmos)



Are they related?
(see page 802.)

the *Mathematische Annalen* affair, van Stigt quotes Brouwer's speculation that Hilbert's motive was anger at Brouwer's declining a position in Göttingen in 1919 and, instead of critically examining this or recalling other documented sources of Hilbert's anger (which would suggest Brouwer was not completely right in his estimation), van Stigt supports the assertion with further speculation. As an indication of just how far from the mark such speculation can get, I might mention van Stigt's remark on Brouwer's fellowship with various young poets: "Brouwer lacked the ability to express his feelings artistically". This certainly runs counter to what every Dutchman I've spoken to has said of Brouwer's correspondence with Adama van Scheltema: Brouwer's prose is uniformly favourably compared with that of the poet. One criticism I've read and must agree with is that van Stigt's psychological analysis of Brouwer is excessive and excessively negative. Indeed, it might be said that van Stigt has done a hatchet job on Brouwer. But, in doing so, he has still managed to paint a more positive image of Brouwer than is generally displayed to mathematicians and the reader can get a first, albeit rough, approximation to Brouwer from van Stigt's book. Thus, with the reservations that the book is only tangentially relevant to the mathematical community and that it must be approached very critically, I refer the reader to it for a faltering first step in getting to know the real Brouwer.

429 South Warwick Avenue
Westmont, IL 60559

There are in this world optimists who feel that any symbol that starts off with an integral sign must necessarily denote something that will have every property that they should like an integral to possess. This of course is quite annoying to us rigorous mathematicians; what is even more annoying is that by doing so they often come up with the right answer.

—E.J. McShane

Bulletin of the American Mathematical Society
v 69, p 611, 1963.

**Answer to Picture Puzzle
(p. 770)**

No, they are not: they are Emile Borel and Armand Borel.

Apropos Two Notes on Notation

Antal E. Fekete

The best thing that has happened to the Stirling numbers during their 400 years of chequered history is Donald Knuth's paper *Two Notes on Notation*. The symbol $\binom{n}{k}$ of von Ettinghausen for the binomial coefficients, introduced in 1826, has been called 'beautiful', suggesting a most successful combination of aesthetics and economy. For precisely the same reasons, Knuth recommends the use of the symbols $\left[\begin{smallmatrix} n \\ k \end{smallmatrix} \right]$ and $\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\}$ to denote the Stirling numbers of the first and second kind, in combination with a *conceptual* definition, leading to the rediscovery of the Law of Reciprocity $\left\{ \begin{smallmatrix} -n \\ -k \end{smallmatrix} \right\} = \left[\begin{smallmatrix} k \\ n \end{smallmatrix} \right]$. The purpose of this note is to explore reciprocity further, and to make additional comments on terminology and notation.

1. BINOMIAL COEFFICIENTS AND RECIPROCITY. Let $\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\}$ denote the number of k -quotient sets of an n -set, and let $\left[\begin{smallmatrix} n \\ k \end{smallmatrix} \right]$ denote the number of k -orbit permutations of an n -set, called the *Stirling numbers* and the *reciprocal Stirling numbers*, respectively. Their earlier names (Stirling numbers of the second and first kind, respectively), have justly been criticized by Donald Knuth in [1] for being historically incorrect and without mnemonic merit. However, his suggestion to call $\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\}$ a "Stirling subset number", and to verbalize the symbol to read " n subset k " would spell conceptual disaster. Reference to subsets in connection with the Stirling numbers must be avoided. Stirling numbers count quotient sets; it is *binomial coefficients* that count subsets. Indeed, the conceptual definition of $\binom{n}{k}$ is the number of k -subsets of an n -set. The lattice of subsets is very different from the lattice of quotient sets of an n -set (for the concept of a quotient set, see [2]). The complementary nature of the two lattices, highlighted by the remarkable formulas for the number of injective/surjective functions:

$$\text{Inj}(k, n) = k! \binom{n}{k}, \quad \text{Surj}(n, k) = k! \left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\}$$

must be carefully preserved. In verbalizing, the best course is to read $\binom{n}{k}$ " n sub k ", and to read $\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\}$ " n quotient k ".

If we wish to extend a two-parameter symbol such as $\binom{n}{k}$ from the natural numbers to integers, we are at the crossroads. We have to choose between the conflicting demands of two conventions. On the one hand, we may have to refer to rows with slope m , in which case the conventions governing plane coordinate systems apply. On the other hand, we may have to do calculations with the infinite matrix represented by the table, in which case the conventions governing matrix calculus apply (i.e., increasing n means moving down, and increasing k means moving right). I go along with the latter (which is also Knuth's choice) but I shall,

all the same, refer to rows with slope m , which then must be interpreted in the traditional coordinate system. This leads to the mild anomaly that the main diagonal has slope -1 while the secondary diagonal has slope 1 . To illustrate the power of this new language I restate the familiar relation between the binomial coefficients and the Fibonacci numbers: the sum of entries in the n th row with slope 1 of Pascal's triangle (Table 2) is equal to the n th Fibonacci number. For the sum of entries in rows with slope m of Pascal's triangle, see [3i], p 91.

The binomial coefficients satisfy the recurrence relation

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

Some recurrence relations admit an extension of the two parameter symbol to negative integral values of n, k ; others do not; still others may admit two essentially different extensions. *The existence of an extension depends on the availability of sloping rows (columns) whose entries are in arithmetic progression.* Indeed, every arithmetic progression can be continued for negative ordinals (note that it is possible to encounter several consecutive zeros without the continuation being identically zero!). The familiar extension of the binomial coefficients to negative values of n owes its existence to the fact that entries in the k th column of Pascal's triangle are in arithmetic progression of order k , yielding the formula

$$\binom{-n}{k} = (-1)^k \binom{n+k-1}{k}$$

and the values (blank spaces stand for 0's)

TABLE 1

.
1	-4	10	-20	35	-56	.
1	-3	6	-10	15	-21	.
1	-2	3	-4	5	-6	.
1	-1	1	-1	1	-1	.
1						
1	1					
1	2	1				
1	3	3	1			
1	4	6	4	1		
1	5	10	10	5	1	
.

Theorem 1. *Entries in the k th column of Table 1*

$$\dots, \binom{k-n}{k}, \dots, \binom{k}{k}, \binom{k+1}{k}, \binom{k+2}{k}, \dots, \binom{k+n}{k}, \dots$$

are in arithmetic progression of order k , and the difference sequence is furnished by the $k-1$ -st column.

Here we have $\binom{n}{k} = 0$ for all $k < 0$, and $\binom{-1}{0} = 1$. However, this is not the only possible extension. Another is based on the fact that entries in the n th row with slope -1 of Pascal's triangle are in arithmetic progression of order n , that can be continued for negative values of k . It yields the *Law of Reciprocity for binomial coefficients*:

$$\binom{1-k}{1-n} = (-1)^{n+k} \binom{n}{k}$$

and the values

TABLE 2

1									
-6	1								
15	-5	1							
-20	10	-4	1						
15	-10	6	-3	1					
-6	5	-4	3	-2	1				
1	-1	1	-1	1	-1	1	0		
							1		
							1	1	
							1	2	1
							1	3	3
							1	4	6
							1	5	10
							1	6	15
							1	7	21
							1	8	28
							1	9	36
							1	10	45
							1	11	55
							1	12	66
							1	13	78
							1	14	91
							1	15	105
							1	16	120
							1	17	136
							1	18	153
							1	19	171
							1	20	190
							1	21	210
							1	22	231
							1	23	253
							1	24	276
							1	25	300
							1	26	325
							1	27	351
							1	28	378
							1	29	406
							1	30	435
							1	31	465
							1	32	496
							1	33	528
							1	34	561
							1	35	595
							1	36	630
							1	37	666
							1	38	703
							1	39	741
							1	40	780
							1	41	820
							1	42	861
							1	43	903
							1	44	946
							1	45	990
							1	46	1035
							1	47	1081
							1	48	1128
							1	49	1176
							1	50	1225
							1	51	1275
							1	52	1326
							1	53	1378
							1	54	1431
							1	55	1485
							1	56	1540
							1	57	1596
							1	58	1653
							1	59	1711
							1	60	1770
							1	61	1830
							1	62	1891
							1	63	1953
							1	64	2016
							1	65	2080
							1	66	2145
							1	67	2211
							1	68	2278
							1	69	2346
							1	70	2415
							1	71	2485
							1	72	2556
							1	73	2627
							1	74	2700
							1	75	2773
							1	76	2847
							1	77	2922
							1	78	2998
							1	79	3075
							1	80	3153
							1	81	3232
							1	82	3312
							1	83	3393
							1	84	3475
							1	85	3558
							1	86	3642
							1	87	3727
							1	88	3813
							1	89	3900
							1	90	3988
							1	91	4077
							1	92	4167
							1	93	4258
							1	94	4350
							1	95	4443
							1	96	4537
							1	97	4632
							1	98	4728
							1	99	4825
							1	100	4923

Theorem 2. Entries in the n th row with slope -1 of Table 2

$$\dots, \binom{k-n}{-n}, \dots, \binom{0}{-k}, \dots, \binom{k}{0}, \binom{k+1}{1}, \binom{k+2}{2}, \dots, \binom{n}{n-k}, \dots$$

are in arithmetic progression of order n , and the difference sequence is furnished by the $n-1$ -st row with slope -1 .

Theorem 2*. Reciprocity for the arithmetic progression of order m in Theorem 2 can also be stated in terms of a determinant of order $m-1$:

$$\begin{pmatrix} 1-k \\ 1-n \end{pmatrix} = \begin{vmatrix} \binom{k+1}{k} & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \binom{n-2}{k} & \binom{n-2}{k+1} & \cdots & 1 & 0 \\ \binom{n-1}{k} & \binom{n-1}{k+1} & \cdots & \binom{n-1}{n-2} & 1 \\ \binom{n}{k} & \binom{n}{k+1} & \cdots & \binom{n}{n-2} & \binom{n}{n-1} \end{vmatrix},$$

The Law of Reciprocity can now be stated in terms of a reflection in the secondary diagonal of Table 2, followed by attaching the signature $(-1)^{n+k}$ to each entry $\binom{n}{k}$. Under this extension $\binom{n}{k} = 0$ for all $n < k$, and $\binom{-1}{0} = 0$. Since there is no way to reconcile the two extensions, we are again at the crossroads. An intelligent choice must be made in order to fix the meaning of the symbol $\binom{n}{k}$ for negative integers. For certain applications, the extension discussed in this note might be preferable. The second natural extension of the binomial coefficients to negative arguments has been noted by Herbert Wilf and Doron Zeilberger [5].

2. STIRLING NUMBERS AND RECIPROCITY. The Stirling numbers and the reciprocal Stirling numbers satisfy the recurrence relations

$$\left\{ \begin{matrix} n \\ k \end{matrix} \right\} = \left\{ \begin{matrix} n-1 \\ k-1 \end{matrix} \right\} + k \left\{ \begin{matrix} n-1 \\ k \end{matrix} \right\} \quad \text{and} \quad \left[\begin{matrix} n \\ k \end{matrix} \right] = \left[\begin{matrix} n-1 \\ k-1 \end{matrix} \right] + (n-1) \left[\begin{matrix} n-1 \\ k \end{matrix} \right]$$

$$\begin{Bmatrix} -k \\ -n \end{Bmatrix} = \begin{bmatrix} n \\ k \end{bmatrix}$$

TABLE 3													
.													
.	1												
.	21	1											
.	175	15	1										
.	735	85	10	1									
.	1624	225	35	6	1								
.	1764	274	50	11	3	1							
.	720	120	24	6	2	1	1						
							1						
							1						
							1	1					
							1	3	1				
							1	7	6	1			
							1	15	25	10	1		
							1	31	90	65	15	1	
							1	63	301	350	140	21	1

$$\dots, \begin{Bmatrix} n-k \\ -k \end{Bmatrix}, \dots, \begin{Bmatrix} 0 \\ -n \end{Bmatrix}, \dots, \begin{Bmatrix} n \\ 0 \end{Bmatrix}, \begin{Bmatrix} n+1 \\ 1 \end{Bmatrix}, \begin{Bmatrix} n+2 \\ 2 \end{Bmatrix}, \dots, \begin{Bmatrix} k \\ k-n \end{Bmatrix}, \dots$$

Theorem 3*. *Reciprocity for the arithmetic progression of order m in Theorem 3 can also be stated in terms of a determinant of order $m - 1$:*

$$\begin{Bmatrix} -k \\ -n \end{Bmatrix} = \begin{vmatrix} \begin{Bmatrix} k+1 \\ k \end{Bmatrix} & 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \begin{Bmatrix} n-2 \\ k \end{Bmatrix} & \begin{Bmatrix} n-2 \\ k+1 \end{Bmatrix} & \cdots & 1 & 0 \\ \begin{Bmatrix} n-1 \\ k \end{Bmatrix} & \begin{Bmatrix} n-1 \\ k+1 \end{Bmatrix} & \cdots & \begin{Bmatrix} n-1 \\ n-2 \end{Bmatrix} & 1 \\ \begin{Bmatrix} n \\ k \end{Bmatrix} & \begin{Bmatrix} n \\ k+1 \end{Bmatrix} & \cdots & \begin{Bmatrix} n \\ n-2 \end{Bmatrix} & \begin{Bmatrix} n \\ n-1 \end{Bmatrix} \end{vmatrix}.$$

[October

signature $(-1)^{n+k}$ to each entry $\begin{Bmatrix} n \\ k \end{Bmatrix}$:

TABLE 3*

$$\begin{array}{cccccccc}
\cdot & & & & & & & \\
\cdot & 1 & & & & & & \\
\cdot & -21 & 1 & & & & & \\
\cdot & 140 & -15 & 1 & & & & \\
\cdot & -350 & 65 & -10 & 1 & & & \\
\cdot & 301 & -90 & 25 & -6 & 1 & & \\
\cdot & -63 & 31 & -15 & 7 & -3 & 1 & \\
\cdot & 1 & -1 & 1 & -1 & 1 & -1 & 1
\end{array}$$

This fact can also be expressed by the formula

$$\sum_k (-1)^{m+k} \begin{Bmatrix} n \\ k \end{Bmatrix} \begin{bmatrix} k \\ m \end{bmatrix} = [n = m]$$

where I have used Iverson's bracket [1].

The proof of Theorem 3 is based on the famous formulas, independently discovered by Appell [6], Jordan [7], and Ward [8], expressing entries in the n th rows with slope -1 as polynomials in k of degree $2n$,

$$\left\{ \begin{matrix} k \\ k-n \end{matrix} \right\} = \left\{ \begin{matrix} k \\ n+1 \end{matrix} \right\} \left\{ \left\{ \begin{matrix} n+1 \\ 1 \end{matrix} \right\} \right\} + \left\{ \begin{matrix} k \\ n+2 \end{matrix} \right\} \left\{ \left\{ \begin{matrix} n+2 \\ 2 \end{matrix} \right\} \right\} + \cdots + \left\{ \begin{matrix} k \\ 2n \end{matrix} \right\} \left\{ \left\{ \begin{matrix} 2n \\ n \end{matrix} \right\} \right\} \cdots \quad (1)$$

$$\begin{bmatrix} k \\ k-n \end{bmatrix} = \begin{pmatrix} k \\ n+1 \end{pmatrix} \begin{bmatrix} n+1 \\ 1 \end{bmatrix} + \begin{pmatrix} k \\ n+2 \end{pmatrix} \begin{bmatrix} n+2 \\ 2 \end{bmatrix} + \dots + \begin{pmatrix} k \\ 2n \end{pmatrix} \begin{bmatrix} 2n \\ n \end{bmatrix} \dots \quad (2)$$

where the coefficients

$$\left\{ \begin{pmatrix} n+i \\ i \end{pmatrix} \right\} = \Delta^{n+i} \begin{pmatrix} 0 \\ -n \end{pmatrix} \quad \text{and} \quad \left\| \begin{pmatrix} n+i \\ i \end{pmatrix} \right\| = \Delta^{n+i} \begin{bmatrix} 0 \\ -n \end{bmatrix}$$

depend on n but not on k , for $i = 1, 2, \dots, n$. Here $\Delta^m \begin{Bmatrix} 0 \\ -n \end{Bmatrix}$ denotes $\Delta^m a_0$ for $a_k = \begin{Bmatrix} k \\ k-n \end{Bmatrix}$. In general, the terms of a sequence b_k are in arithmetic progression of order n if, and only if, b_k is a polynomial in k of degree n , in which case the n th difference sequence is constant and is equal to $n!$ -times the coefficient of k^n . It follows that the $2n$ th difference sequence of the progression a_n is constant and equal to $\left\{ \begin{Bmatrix} 2n \\ n \end{Bmatrix} \right\}$ and $\llbracket 2n \rrbracket$, respectively. Knuth gives reference to several proofs of the fact that (1) and (2) are polynomials in k of degree $2n$. In my presentation it is clear that (1) and (2) are special cases of Newton's first formula expressing the subsequent terms of the arithmetic progression of Theorem 3. This remark makes further proof unnecessary. However, I shall furnish yet another proof which has

the merit of making these formulas conceptually transparent. The symbol $\left\{\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\} \right\}$ is called a *second order Stirling number*, defined as the number of k -quotient sets of an n -set having at least two elements in each equivalence class; and the symbol $\left[\left[\begin{smallmatrix} n \\ k \end{smallmatrix} \right] \right]$ is called a *second order reciprocal Stirling number* defined as the number of k -orbit permutations of an n -set having at least two elements in each orbit. In other words, the permutations have no fixed points and therefore are *derangements*, thus $\left[\left[\begin{smallmatrix} n \\ k \end{smallmatrix} \right] \right]$ is the number of k -orbit derangements of an n -set.

The proof of (1) is based on a survey of quotient sets of a k -set with $k - n$ equivalence classes; these clearly have at least $k - 2n$, but no more than $k - n - 1$, single-element classes. The number of those having exactly $k - n - i$ single-element classes is obviously

$$\binom{k}{n+i} \left\{\left\{ \begin{smallmatrix} n+i \\ i \end{smallmatrix} \right\} \right\}.$$

If we add up these numbers for $i = 1, 2, \dots, n$, then we have accounted for every quotient set under survey exactly once, proving (1); (2) is proved *mutatis mutandis*. We have:

$$\begin{aligned} \left\{\left\{ \begin{smallmatrix} n \\ 1 \end{smallmatrix} \right\} \right\} &= 1, \left[\left[\begin{smallmatrix} n \\ 1 \end{smallmatrix} \right] \right] = (n-1)! \quad \text{for } n > 1; \left\{\left\{ \begin{smallmatrix} 1 \\ 1 \end{smallmatrix} \right\} \right\} = 0 = \left[\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix} \right] \right] \\ \text{and } \left\{\left\{ \begin{smallmatrix} 2n \\ n \end{smallmatrix} \right\} \right\} &= 1.3.5 \dots (2n-1) = \left[\left[\begin{smallmatrix} 2n \\ n \end{smallmatrix} \right] \right] \end{aligned}$$

because in the latter case each equivalence class (orbit) has exactly 2 elements. This completes the proof of Theorem 3. We also have

$$\left\{\left\{ \begin{smallmatrix} 2n \\ n+k \end{smallmatrix} \right\} \right\} = 0 = \left[\left[\begin{smallmatrix} 2n \\ n+k \end{smallmatrix} \right] \right] \quad \text{and} \quad \left\{\left\{ \begin{smallmatrix} 2n+1 \\ n+k \end{smallmatrix} \right\} \right\} = 0 = \left[\left[\begin{smallmatrix} 2n+1 \\ n+k \end{smallmatrix} \right] \right],$$

for $k = 1, 2, 3, \dots$

because if a quotient set of a $2n$ -set, or a $2n + 1$ -set, has more than n equivalence classes (orbits), then at least one of them must be a single-element class (orbit).

The recurrence relations

$$\left\{\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\} \right\} = (n-1) \left\{\left\{ \begin{smallmatrix} n-2 \\ k-1 \end{smallmatrix} \right\} \right\} + k \left\{\left\{ \begin{smallmatrix} n-1 \\ k \end{smallmatrix} \right\} \right\}$$

and

$$\left[\left[\begin{smallmatrix} n \\ k \end{smallmatrix} \right] \right] = (n-1) \left[\left[\begin{smallmatrix} n-2 \\ k-1 \end{smallmatrix} \right] \right] + (n-1) \left[\left[\begin{smallmatrix} n-1 \\ k \end{smallmatrix} \right] \right]$$

can be proved by a simple combinatorial argument, not reproduced here, and they yield the numbers

TABLE FOR $\left\{\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\} \right\}$					TABLE FOR $\left[\left[\begin{smallmatrix} n \\ k \end{smallmatrix} \right] \right]$				
0					0				
1	0				1	0			
1	0	0			2	0	0		
1	3	0	0		6	3	0	0	
1	10	0	0	0	24	20	0	0	0
1	25	15	0	0	120	130	15	0	0
1	56	105	0	0	720	924	210	0	0
1	119	490	105	0	5040	7308	2380	105	0
1	246	1918	1260	0	40320	64224	26432	2520	0
.

Note that the coefficients in Formulas (1) and (2) are exactly those in the rows of these tables with slope -1 . More extensive tables are found in [3ii] pp. 57, 98 and in [4] pp 152, 172. Second order Stirling numbers can also be expressed explicitly in terms of the Stirling numbers, thanks to Newton's second formula:

$$\left\{\left\{\begin{matrix} n \\ k \end{matrix}\right\}\right\} = \sum_i (-1)^i \binom{n}{n-i} \left\{\begin{matrix} n-i \\ k-i \end{matrix}\right\} \quad \text{and} \quad \left[\left[\begin{matrix} n \\ k \end{matrix}\right]\right] = \sum_i (-1)^i \binom{n}{n-i} \left[\begin{matrix} n-i \\ k-i \end{matrix}\right].$$

More generally, a Stirling number of order r is the number of k -quotient sets of an n -set having at least r elements in each equivalence class, and a reciprocal Stirling number of order r is the number of k -orbit permutations of an n -set having at least r elements in each orbit. I recommend the notation

$$\left\{\left\{\begin{matrix} n \\ k \end{matrix}\right\}\right\}_r, \left[\left[\begin{matrix} n \\ k \end{matrix}\right]\right]_r \quad \text{with} \quad \left\{\left\{\begin{matrix} n \\ k \end{matrix}\right\}\right\}_2 = \left\{\left\{\begin{matrix} n \\ k \end{matrix}\right\}\right\}, \left\{\left\{\begin{matrix} n \\ k \end{matrix}\right\}\right\}_1 = \left\{\begin{matrix} n \\ k \end{matrix}\right\}$$

and

$$\left[\left[\begin{matrix} n \\ k \end{matrix}\right]\right]_2 = \left[\left[\begin{matrix} n \\ k \end{matrix}\right]\right], \left[\left[\begin{matrix} n \\ k \end{matrix}\right]\right]_1 = \left[\begin{matrix} n \\ k \end{matrix}\right]$$

to denote these numbers, which satisfy the recurrence relations

$$\begin{aligned} \left\{\left\{\begin{matrix} n \\ k \end{matrix}\right\}\right\}_r &= \binom{n-1}{r-1} \left\{\left\{\begin{matrix} n-r \\ k-1 \end{matrix}\right\}\right\}_r + k \left\{\left\{\begin{matrix} n-1 \\ k \end{matrix}\right\}\right\}_r; \left[\left[\begin{matrix} n \\ k \end{matrix}\right]\right]_r \\ &= (n-1)(n-2) \cdots (n-r+1) \left[\left[\begin{matrix} n-r \\ k-1 \end{matrix}\right]\right]_r + (n-1) \left[\left[\begin{matrix} n-1 \\ k \end{matrix}\right]\right]_r. \end{aligned}$$

Related concepts are that of an r -Stirling number [9], and that of a weighted Stirling number [10].

3. CONCLUSION. The main result of this note is Theorem 3*, establishing an explicit formula for the reciprocals of Stirling numbers. This determinant formula may be used to discover the yet unknown reciprocals of certain other numbers dependent on two integral parameters, for example, the Lah numbers which, like the binomial coefficients, are self-reciprocal, and the idempotent numbers which like the Stirling numbers, are not—but their reciprocals have an interesting combinatorial interpretation. (For the definition of the Lah numbers and the idempotent numbers, see [3]).

ACKNOWLEDGMENTS. I am indebted to Donald Knuth and Herbert Wilf for valuable comments they have made on the first draft, and also for calling my attention to important references.

REFERENCES

1. Donald E. Knuth, Two Notes on Notation, *Am. Math. Monthly*, vol. 99 no. 5 (May, 1992) 403–422.
2. Roger Godement, *Algebra*, Boston, 1968.
3. Louis Comtet, *Analyse Combinatoire*, Tomes i et ii, Paris, 1970 (English version. Advanced Combinatorics, Dordrecht, Holland, 1974).
4. Charles Jordan, *Calculus of Finite Differences*, 2nd edition reprinted by Chelsea, New York, 1960.
5. Herbert Wilf and Doron Zeilberger, Rational Functions Certify Combinatorial Identities, *Journ. AMS* (1990) 147–158.
6. P. Appell, Développement en série entière de $(1+ax)^{1/x}$, *Arch. der Math. und Phys.*, vol. 65 (1880), 171–175.

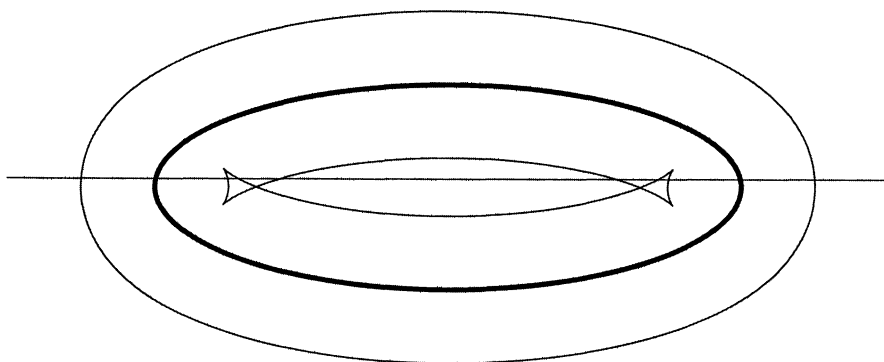
7. Charles Jordan, On Stirling Numbers, *Tôhoku Math. Journ.*, vol. 37 (1933), 254–278.
8. M. Ward, The Representation of Stirling's Numbers and Stirling's Polynomials as a Sum of Factorials, *Am. Journ. of Math.*, vol. 56 (1934), 87–95.
9. Andrei Z. Broder, The r -Stirling Numbers, *Discrete Maths*, 49 (1984), no. 3, 241–259.
10. L. Carlitz, Weighted Stirling Numbers of the First and Second Kind, Parts I and II, *Fibonacci Quart.*, vol. 18 (1980), no. 2, 147–162 and no 3, p; 242–257.

Department of Mathematics and Statistics
 Memorial University of Newfoundland
 St. John's, Canada A1C 5S7
 fekete@riemann.math.mun.ca

PC Curves

I write to correct a remark made by Thomas Banchoff and Peter Giblin in their article “On the geometry of piecewise circular curves” (May 1994, pp. 403–416).

On page 405, they claim that an exterior offset of an ellipse is an algebraic curve of degree 4. In fact, offsetting an ellipse by a distance d produces a single algebraic curve of degree 8. It is only when we restrict from the complex numbers to the reals that this curve breaks up into separate interior and exterior components.



Looking at examples in which the offset distance d is small, one might be fooled into thinking that the interior and exterior offsets were separate curves, each of degree 4. But in the example above, where a straight line intersects an offset at eight distinct, real points, six of those points lie on the interior component.

Dr. Lyle Ramshaw
 Digital Equipment Corporation
 Systems Research Center
 130 Lytton Avenue
 Palo Alto, CA 94301

NOTES

Edited by: John Duncan

The Coin Exchange Problem for Arithmetic Progressions

Amitabha Tripathi

We consider a problem which, in the general case, dates back to Frobenius, and can be thought of as exchanging coins of arbitrary denomination with an infinite supply of coins of certain fixed denominations.

Hence, we are given integers a_0, \dots, a_{k-1} , and N , and we seek nonnegative integers x_0, \dots, x_{k-1} such that

$$a_0x_0 + \dots + a_{k-1}x_{k-1} = N. \quad (1)$$

In this note, we will consider the case where the given coin denominations a_j are in arithmetic progression. We write $a_j = a + (j-1)d$, with $(a, d) = 1$, for $1 \leq j \leq k$. For a fixed value of $k \geq 2$, we denote by $g(a, d; k)$ (respectively, $n(a, d; k)$) the largest (respectively, number of) N such that

$$a \left(\sum_{j=0}^{k-1} x_j \right) + d \left(\sum_{j=0}^{k-1} jx_j \right) = N \quad (2)$$

has no solution in nonnegative integers.

We present a simple argument that results in a formula for $g(a, d; k)$ and $n(a, d; k)$. A variation of the formula for $g(a, 1; k)$ is due to [Bra42] while that for $n(a, 1; k)$ is due to [NW72]. [Rob56] generalized this to obtain $g(a, d; k)$, later simplified by [Bat58], and [Gra73] obtained the result for $n(a, d; k)$.

Let \mathcal{C} denote a non zero residue class modulo a , and let $m_{\mathcal{C}}$ denote the least positive integer of the form $a(\sum_{j=0}^{k-1} x_j) + d(\sum_{j=0}^{k-1} jx_j)$ in \mathcal{C} . It is well known that both $g(a, d; k)$ and $n(a, d; k)$ can be readily derived from these minima. A derivation of these may be found in [Tri89], but we present it here for the sake of completeness.

Lemma 1.

- (i) $g(a, d; k) = \max_{\mathcal{C}} m_{\mathcal{C}} - a$, the maximum taken over all nonzero classes \mathcal{C} .
- (ii) $n(a, d; k) = 1/a \sum_{\mathcal{C}} m_{\mathcal{C}} - (a-1)/2$, the sum taken over all nonzero classes \mathcal{C} .

Proof: (i) Since $\max_{\mathcal{C}} m_{\mathcal{C}}$ and $\max_{\mathcal{C}} m_{\mathcal{C}} - a$ are in the same residue class modulo a , and $\max_{\mathcal{C}} m_{\mathcal{C}}$ is the least positive integer of the form (2) in its class, $g(a, d; k) \geq \max_{\mathcal{C}} m_{\mathcal{C}} - a$. On the other hand, if $N > \max_{\mathcal{C}} m_{\mathcal{C}} - a$, then $N \geq \max_{\mathcal{C}} m_{\mathcal{C}}$ for each class \mathcal{C} , so that N is of the given form.

(ii) The number of positive integers in class \mathcal{C} which cannot be represented in the form given by (2) is $[m_{\mathcal{C}}/a]$, since $m_{\mathcal{C}}$ is the least representable integer in

class \mathcal{C} . Hence, the total number of nonrepresentable integers is

$$\sum_{\mathcal{C}} \left\lceil \frac{m_{\mathcal{C}}}{a} \right\rceil = \sum_{\mathcal{C}} \frac{m_{\mathcal{C}}}{a} - \sum_{i=0}^{a-1} \frac{i}{a} = \frac{1}{a} \sum_{\mathcal{C}} m_{\mathcal{C}} - \frac{a-1}{2}. \quad \square$$

We note that the lemma together with its proof carries over to the general case given by (1), with g and n being defined analogously.

Lemma 2. *For each y , $1 \leq y \leq a-1$, the least positive integer of the form given by (2) in the class $dy \bmod a$ is given by*

$$a \left(1 + \left\lceil \frac{y-1}{k-1} \right\rceil \right) + dy.$$

Proof: Each nonzero class \mathcal{C} modulo a determines a unique y , with $1 \leq y \leq a-1$, such that $dy \in \mathcal{C}$. This clearly is the smallest multiple of d in the class \mathcal{C} . For this choice of y , we wish to minimize $\sum_{j=0}^{k-1} x_j$ in order to determine the smallest member of \mathcal{C} . With $y = (k-1)q + r$, $0 \leq r \leq k-2$, we may choose $x_{k-1} = q$, with $x_r = 1$, other $x_j = 0$ provided $r \neq 0$ but with all other $x_j = 0$ if $r = 0$. Thus, the minimum value for $\sum_{j=0}^{k-1} x_j$ is $q+1$ if $r \neq 0$ and q if $r = 0$. This may be written more briefly as

$$\left(1 + \left\lceil \frac{y-1}{k-1} \right\rceil \right),$$

so that

$$m_{\mathcal{C}} = a \left(1 + \left\lceil \frac{y-1}{k-1} \right\rceil \right) + dy, \quad \text{where } dy \in \mathcal{C}. \quad \square$$

Theorem 1.

- (i) $g(a, d; k) = a[(a-2)/(k-1)] + d(a-1)$;
- (ii) $n(a, d; k) = \frac{1}{2}(a+t)(1 + [(a-2)/(k-1)]) + \frac{1}{2}(a-1)(d-1)$, where t is the smallest nonnegative integer such that $a-2 \equiv t \pmod{k-1}$.

Proof: The theorem is an easy consequence of the two lemmas.

$$\begin{aligned} \text{(i)} \quad g(a, d; k) &= \max_{\mathcal{C}} m_{\mathcal{C}} - a \\ &= \max_{1 \leq y \leq a-1} a \left\lceil \frac{y-1}{k-1} \right\rceil + dy \\ &= a \left\lceil \frac{a-2}{k-1} \right\rceil + d(a-1). \end{aligned}$$

(ii)

$$\begin{aligned}
n(a, d; k) &= \frac{1}{a} \sum_{\mathcal{L}} m_{\mathcal{L}} - \frac{a-1}{2} \\
&= \frac{1}{a} \sum_{y=1}^{a-1} a \left(1 + \left\lfloor \frac{y-1}{k-1} \right\rfloor \right) + dy - \frac{a-1}{2} \\
&= \sum_{y=0}^{a-2} \left(1 + \left\lfloor \frac{y}{k-1} \right\rfloor \right) + \frac{d(a-1)}{2} - \frac{a-1}{2} \\
&= \sum_{y=0}^{[(a-2)/(k-1)](k-1)-1} \left(1 + \left\lfloor \frac{y}{k-1} \right\rfloor \right) \\
&\quad + \sum_{y=[(a-2/k-1)](k-1)}^{a-2} \left(1 + \left\lfloor \frac{y}{k-1} \right\rfloor \right) + \frac{(a-1)(d-1)}{2} \\
&= (k-1) \left(1 + 2 + \cdots + \left\lfloor \frac{a-2}{k-1} \right\rfloor \right) \\
&\quad + (1+t) \left(1 + \left\lfloor \frac{a-2}{k-1} \right\rfloor \right) + \frac{(a-1)(d-1)}{2}, \\
&\quad \text{where } t \equiv a-2 \pmod{(k-1)} \\
&= \frac{1}{2} \left(1 + \left\lfloor \frac{a-2}{k-1} \right\rfloor \right) \left((k-1) \left\lfloor \frac{a-2}{k-1} \right\rfloor + t + 2 + t \right) \\
&\quad + \frac{(a-1)(d-1)}{2} \\
&= \frac{1}{2} \left(1 + \left\lfloor \frac{a-2}{k-1} \right\rfloor \right) (a+t) + \frac{(a-1)(d-1)}{2}, \\
&\quad \text{where } t \equiv a-2 \pmod{(k-1)}. \quad \square
\end{aligned}$$

ACKNOWLEDGMENT. The author wishes to thank Preeti Nigam and the referee for their helpful suggestions.

REFERENCES

-
- [Bat58] P. T. Bateman. Remark on a recent note on linear forms. *AMM*, 65:517–518, 1958.
[Bra42] A. Brauer. On a problem of partitions. *AJM*, 64:299–312, 1942.
[Gra73] D. D. Grant. On linear forms whose coefficients are in arithmetic progression. *Israel Journal of Mathematics*, 15:204–209, 1973.
[NW72] A. Nijenhuis and H. S. Wilf. Representations of integers by linear forms in nonnegative integers. *Journal of Number Theory*, 4:98–106, 1972.
[Rob56] J. B. Roberts. Note on linear forms. *Proceedings of the American Mathematical Society*, 7:465–469, 1956.
[Tri89] A. Tripathi. *Topics in number theory*. Ph.D. thesis, Dept. of Mathematics, SUNY, Buffalo, 1989.

Department of Mathematics
Indian Institute of Technology
New Delhi 110016, India
atripath@netearth.iitd.ernet.in

Congruence of Triangles

Leonard Gillman

“They seem to have forgotten all their high school geometry.” This could be almost any college teacher speaking. Can part of the problem be that the course has become dull? I suggest injecting some mathematical spirit to freshen it up.

I got to thinking about this recently on reading Daniel Hirschhorn’s interesting article [1] about “SSA”, the ambiguous case in the congruence of triangles. It turns out that if the angle is the one opposite the larger side, then there is no ambiguity but a theorem—in Hirschhorn’s clever notation, the SsA theorem. Hirschhorn laments that although the result was published at least a dozen years ago [2, 3], it has not yet made its way into the textbooks.

I too lament the fact that the theorem has not caught on. One reason may be that SsA is an outsider, a role accentuated by the fact that the classical SSS, SAS, and ASA appear in most of today’s textbooks only as postulates.

Another possible reason is the proofs. Even though the proof in [3] is concise, it is still indirect, and students tend to be uncomfortable with indirect proofs. Hirschhorn supplies a direct proof, but it is much more complicated than the other one and in fact is a veritable tour de force of diagrams and notation.

I propose introducing the notation of a representative of an equivalence class. From this point of view, congruence conditions *determine* a triangle (that is, except for its location and orientation). For example, two sides and the included angle determine a triangle. The four congruence proofs are then put on an equal footing. Figures 1–4 are proofs without words—in fact, without letters.

Here are some words. In Figure 1, one side (drawn heavy) determines two of the vertices, and the intersection of two appropriate arcs then locates the third vertex. In Figure 2, the remaining two vertices are the endpoints of the two sides forming the given angle. In Figure 3, the third vertex is the intersection of the rays forming the given angles with the given side.

In Figure 4, the given shorter side is drawn heavy. The argument rests on the fact that the longer of two segments from a point to a line cuts off the greater distance from the perpendicular. In the triangle on the right, the given angle is an

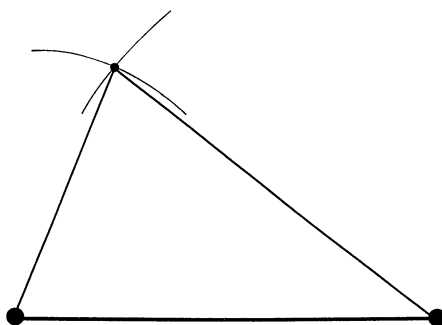


Figure 1. SSS

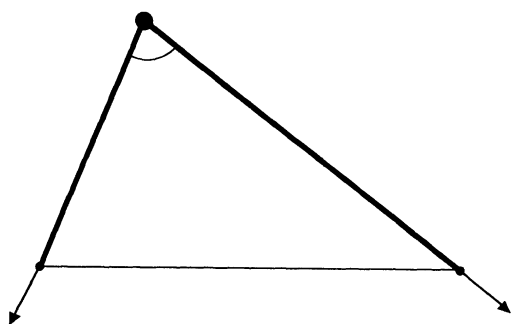


Figure 2. SAS

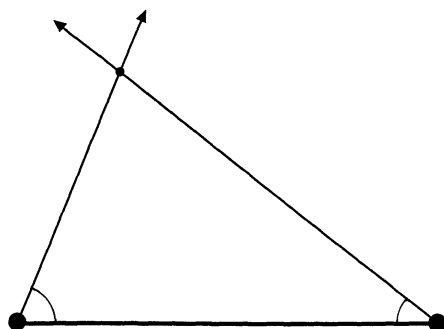


Figure 3. ASA

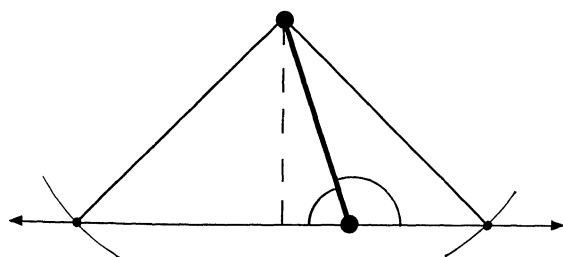


Figure 4. SsA

obtuse or a right angle; on the left, it is acute. In each case there is exactly one location for the third vertex.

REFERENCES

1. Daniel B. Hirschhorn, Why Is the SsA Triangle-Congruence Theorem Not Included in Textbooks?, *Math. Teacher* 83 (1990), 358–361.
2. Bonnie H. Litwiller and David R. Duncan, SSA: When Does It Yield Triangle Congruence?, *Math. Teacher* 74 (1981) 106–108.
3. Shraga Yeshurun and David C. Kay, An Improvement on SSA Congruence for Geometry and Trigonometry, *Math. Teacher* 76 (1983) 364–367.

1606 The High Road
Austin, TX 78746
len@math.utexas.edu

A Short Elementary Proof of the Mohr-Mascheroni Theorem

Norbert Hungerbühler

1. INTRODUCTION. In 1797 Lorenzo Mascheroni surprised the mathematical world with the theorem that every geometric construction that can be carried out by compasses and ruler may be done without ruler (see [4]). It turned out later that Georg Mohr proved this theorem in 1672 already (see [6]). The proofs given by Mohr and Mascheroni are quite complicated. Later easier proofs have been developed (See [3] or [5]). Furthermore the proof could be simplified by means of the circular inversion (see [1] or [2]). Here we give a very short and direct proof for the theorem that does not appeal to inversion.

2. THE MOHR-MASCHERONI THEOREM

Theorem. *Every geometric construction carried out by compasses and ruler can be done without ruler.*

Proof: We have to prove that the following three fundamental constructions are possible to carry out with compasses alone.

1. Points of intersection of two circles given by its centers and radii.
2. Points of intersection of a circle (given by center and radius) and a straight line (given by two points).
3. Point of intersection of two straight lines each of them given by two points.

There is nothing to prove for the intersection of two circles, so let us consider

2.1. Points of intersection of a circle and a straight line. Here we have to distinguish two cases:

1. The straight line misses the center of the circle.
2. The straight line passes through the center of the circle.

The first case is covered by the following construction:

Construction 1. If the straight line g is given by the points P_1 and P_2 , we reflect the center M of the given circle K with respect to g as Figure 1 indicates. Then we find the two points of intersection $\{X, Y\} = K \cap g$ as the points of intersection of K and the reflected circle K' .

Before we are able to attack the second case, we need to have a construction which allows to bisect a segment AB without ruler. This can be done as follows:

Construction 2. Let K_1 be the circle through B with center A and K_2 the circle through A with center B with $K_1 \cap K_2 = \{C, D\}$ (see Figure 2). We then find a point E as the intersection of K_2 and the circle K_3 through D around C . Note that B is the bisection point of AE . Let F and G be the points of intersection of K_1 and the circle K_4 through A around E . Then we get the bisection point M of

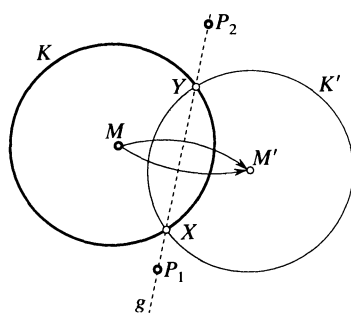


Figure 1

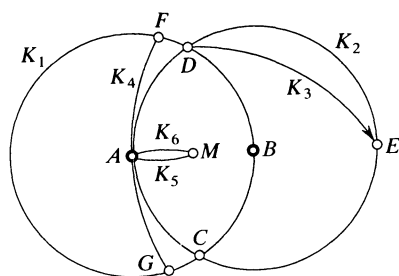


Figure 2

AB as intersection of the circles K_5 through A around F and K_6 through A around G .

The correctness of the construction is evident: Note that the triangles FAM and EFA are similar with proportion $1:2$.

Remark 1. Note that AE has double the length of the segment AB !

Now we construct the points of intersection X and Y of a circle K with center M and a straight line MP :

Construction 3. Let A be an arbitrary point on K and $K \cap AP = \{A, B\}$ (see Figure 3). B is constructed according to construction 1. Let K_1 be a circle through A and B with radius larger than the radius R of K and M_1 the center of K_1 . Now we construct a segment CD with endpoints on K_1 and length $2R$ (see Remark 1). Then we obtain P' as the intersection of CD and the circle K_2 through P around M_1 according to construction 1. Let M_3 be the bisection point of CD (see construction 2) and K_3 the circle around M_3 through C . Let E be a point of K_3 with $P'E = PB$. Now X and Y lie on K and $BX = EC$ and $BY = ED$.

The correctness of the construction can be verified as follows: Note that $PX \cdot PY = PA \cdot PB = P'C \cdot P'D$ by applying Euler's Theorem on intersecting secants, once for K and then for K_1 . Hence the sets of points P, Y, M, X, B and P', D, M_3, C, E are congruent by construction. Thus in fact X and Y are obtained as described.

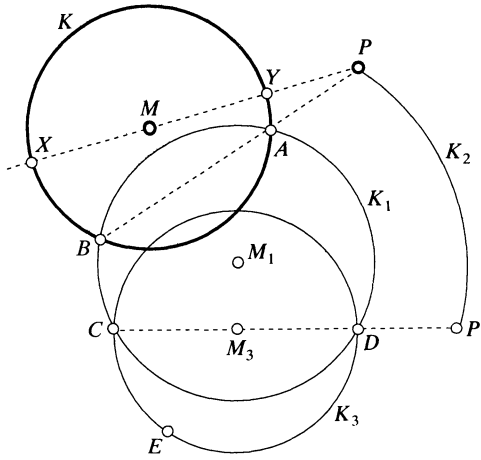


Figure 3

2.2. Point of intersection of two straight lines. Here we first need the following construction with compasses alone which allows to construct the footpoint L of the perpendicular through a point Q on a straight line P_1P_2 :

Construction 4. Just reflect Q with respect to P_1P_2 (see Figure 4). If Q' is the reflected point, we find L as the bisection point of QQ' by Construction 2.

Let us now analyze the situation of two straight lines P_1P_2 and Q_1Q_2 intersecting in S (see Figure 5):

Let L be the footpoint of the perpendicular through Q_1 on P_1P_2 and N be the footpoint of the perpendicular through L on Q_1Q_2 . Both L and N are obtained by Construction 4. Hence we have the relation

$$(Q_1L)^2 = Q_1N \cdot Q_1S.$$

The idea is now to construct the length l of Q_1S since then we find S as intersection of Q_1Q_2 and a circle with center Q_1 and radius l (see Construction 3) and we are through! In fact l is obtained as follows:

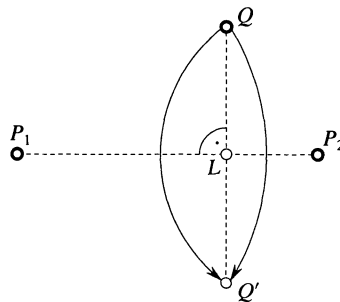


Figure 4

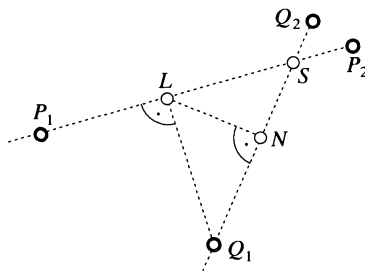


Figure 5

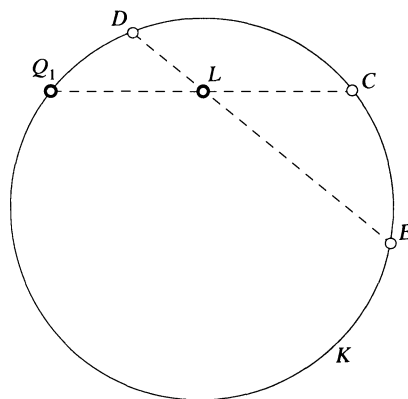


Figure 6

Construction 5. First we double Q_1L ($Q_1C = 2Q_1L$) according to Remark 1 (see Figure 6). Let K be an arbitrary (but large enough) circle through Q_1 and C and let D be a point of K with $LD = Q_1N$. Further let E denote the intersection of LD and K (see construction 1). Then LE has length l since we have $(Q_1L)^2 = Q_1L \cdot LC = LD \cdot LE = Q_1N \cdot LE$ by Euler's theorem on intersecting chords in a circle.

REFERENCES

1. H. Dörrie, *100 Famous Problems in Mathematics*, München: Oldenbourg, 1960.
2. H. Eves, *A Survey of Geometry*, Allyn and Bacon, 1972.
3. R. Honsberger, *Ingenuity in Mathematics*, Washington: Math. Association of America, 1970, ib. 7 (New Mathematical Library, 23).
4. A. N. Kostovskii, *Geometrical Constructions Using Compasses Only*, Blaisdell Publication Company, 1961.
5. Lorenzo Mascheroni, *Geometrie du compas*, Coubron: Monom, 1980.
6. Georg, Mohr, *Euclides danicus*, Amsterdam: Van Velsen, 1672 (Kobenhavn: 1928).

Mathematik Department
 ETH Zürich
 CH-8092, Zürich
 buhler@math.ethz.ch

THE AUTHORS

MARIA TERRELL graduated from Albertus Magnus College and received her Ph.D. at the University of Virginia. Her research focuses on geometry, its history, and its relation to vision. She is currently developing an introductory level geometry course, with support from an NSF curriculum development grant.

ROBERT TERRELL graduated from North Carolina State University, and received his Ph.D. from the University of Virginia. He was an engineer for a time in industry, has patents for machinery, and currently enjoys teaching mathematics to engineering students.

The Terrells were students of E. E. Floyd, who was not only a great mathematician, but a superb teacher.

JAMES BULLOCK presently holds a position as Assistant Professor of Physiology at the University of Missouri-Columbia School of Medicine. His academic credentials are in the field of biology, having earned a baccalaureate degree in that subject from the Illinois Institute of Technology and been awarded the Ph.D. degree in biophysics and theoretical biology by the University of Chicago. The phenomenon that has captured most of his attention is the transport of ions across cell membranes. He points out that biologists have long recognized transport processes of all sorts as crucial to living organisms, and have relied heavily upon principles of physical science and the corresponding mathematics in analyzing questions of function. That mathematics is ubiquitous in the literature of physiology but scarce in its classrooms would have been sufficient to cause him some apprehension. The concern expressed in his article is, in the parlance of his fellow biologists, that the lesion may be more widespread.

GORO SHIMURA is self-taught to a good extent, since he attended neither graduate school nor kindergarten, though he managed to graduate from the University of Tokyo in 1952. Princeton University has been liberal enough to permit him to teach since 1962. His youthful aspiration was to become a fashion designer, but he reconciled himself to living in a less competitive world of mathematics, and testing his meager knowledge of low-dimensional topology by designing his wife's dress occasionally.

BARBARA L. OSOFSKY received her BA and MA from Cornell University, and her PHD from Rutgers University where she has been teaching ever since. Her primary research interest is modules over non-commutative rings, with excursions into commutative rings and algebraic applications of working with infinite cardinals. She has been on the editorial boards of the Proceedings of the AMS and the Journal of Algebra. For health and recreation, she has swum in a great many swimming pools all over the USA and abroad.

SÁNDOR SZABÓ'S main mathematical interest is in algebra and its application. In particular, he has worked on factors of groups related to Hajós' theorem. Recently some of his results played a role in the construction of a tiling of 10-dimensional Euclidean space by congruent parallel cubes such that no pairs have a complete 9-dimensional facet. (This refutes a conjecture Keller made in 1930.) Based at the Technical University of Budapest, he frequently indulges his desire to see the world, serving as a visiting professor at the University of Dundee, University of California at Davis, University of the Pacific, and presently at the University of Bahrain.

SHERMAN STEIN of the University of California at Davis has recently been applying algebra to geometric problems and, with Dean Hickerson, investigating combinatorial aspects of algebraic identities. In addition he has maintained a strong involvement in pedagogy. One fruit of this interest is a 3-year series of high school texts written with Cal Crabill and Don Chakerian and published by WINGS which emphasizes discovery and communication.

ANTAL E. FEKETE is a graduate of the Eötvös Loránd University of Budapest (1955) and has an M.Sc. from the University of Ottawa (1958). He has been on the faculty of Memorial University of Newfoundland since 1958, with several one-year absences while on tour-of-duty elsewhere (Columbia University, 1960; Trinity College, Dublin, 1965; Acadia University, 1970; Princeton University, 1975; American Institute for Economic Research, Great Barrington, Mass., 1983; U.S. House of Representatives, 1989.) He is the author of *Real Linear Algebra* (Marcel Dekker, N.Y., 1985) and several *Apropos* articles, which he hopes to develop into a mathematical literary *genre*.

FOLKE ERIKSSON studied at the University of Uppsala, Sweden, and for a half year in Copenhagen for Werner Fenchel. He has been teaching at Chalmers University of Technology since 1956, except for one year at Western Washington in Bellingham. He is now retiring. He has written a textbook (Swedish) on Calculus in Several Variables. Questions related to Solid Angles and the Law of Sines in dimension > 2 have been his main mathematical interest in the last years.

CRAIG SMORYŃSKI received his Ph.D. at the University of Illinois at Chicago Circle in 1973. His research interests include logic and, more recently, the history of mathematics. He has written articles on Gödel's Theorems for the *Handbook of Mathematical Logic* and the *Handbook of Philosophical Logic*. His other published works include two books for Springer-Verlag, *Self-Reference and Modal Logic* and *Logical Number Theory I; An Introduction*.

Buffon Noodles

Immediately after my article Buffon Noodles appeared in the June/July, '94 issue of this Monthly I received a reference by electronic mail to a very interesting paper by J.F.C. Kingman which, along with a number of lively discussion papers and other references, will clearly be of interest to readers thinking about alternative approaches and refinements of the noodle problem. The precise reference is:

Kingman, J.F.C. (1982), The Thrown String, *J.R. Statist. Soc. B* **44**(2), 109–138

Thank you Colin Mallows at AT & T Bell Labs for the e-mail message.

*Professor Edward C. Waymire
Department of Mathematics
Oregon State University
Kidder Hall 368
Corvallis, OR 97331-4605*

PROBLEMS AND SOLUTIONS

Edited by:
Richard T. Bumby, Fred Kochman and Douglas B. West

Proposed problems should be sent to the MONTHLY PROBLEMS address given on the inside front cover. Please include solutions and relevant references. Three copies of all items needed to evaluate the problem should be sent.

Solutions of published problems should arrive before March 31, 1995 at the MONTHLY PROBLEMS address given on the inside front cover. If possible, solutions should be typed with double spacing. Two copies suffice. Several solutions may be mailed together, but they should be on separate sheets of paper. The problem number and the solver's name and mailing address should appear on each solution. A mailing label should be included if an acknowledgment is desired.

The published solution is likely to be based on a solution that is complete and correct. Additional information, such as references to other appearances of the problem or its solution, is also welcome.

An asterisk () after the number of a problem, or part of a problem, indicates that no solution is currently available.*

PROBLEMS

10403. *Proposed by David Doster, Choate Rosemary Hall, Wallingford, CT.*

Define a sequence $\langle y_n \rangle$ recursively by $y_0 = 1$, $y_1 = 3$ and

$$y_{n+1} = (2n + 3)y_n - 2ny_{n-1} + 8n$$

for $n \geq 1$. Find an asymptotic formula for y_n .

10404. *Proposed by Behzad Djafari Rouhani, Shahid Beheshti University and Islamic Azad University, Tehran, Iran.*

Let x_1, x_2, \dots be a sequence of real numbers such that

$$|x_i - x_j| \geq |x_{i+1} - x_{j+1}|$$

for all positive integers i, j with $|i - j| \leq 2$. Prove that $\langle x_n/n \rangle$ converges to a finite limit as $n \rightarrow \infty$.

10405. Proposed by Herbert Gülicher, Westfälische Wilhelms-Universität, Münster, Germany.

Let $A_1A_2A_3A_4A_5A_6$ be a hexagon circumscribed about a conic, and form the intersections $P_i = A_iA_{i+2} \cap A_{i+1}A_{i+3}$ ($i = 1, \dots, 6$, all indices mod 6). Show that the P_i are the vertices of a hexagon inscribed in a conic.

10406. Proposed by David C. Fisher, University of Colorado, Denver, CO, Karen L. Collins, Wesleyan University, Middleton, CT, and Lucia B. Krompart, Rochester, MI.

Show that a path on an m by n square grid which starts at the northwest corner, goes through each point exactly once, and ends at the southeast corner divides the grid into two equal halves: (a) those regions opening north or east; and (b) those regions opening south or west.

10407. Proposed by Roy Mathias, College of William & Mary, Williamsburg, VA.

Let $\lambda_1, \dots, \lambda_{n+1}$ and μ_1, \dots, μ_n be $2n + 1$ given real numbers such that

$$\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \mu_2 \leq \dots \leq \lambda_n \leq \mu_n \leq \lambda_{n+1}$$

$$\mu_1 < \mu_2 < \dots < \mu_n.$$

Show that

$$\sum_{j=1}^n \frac{\prod \left\{ \lambda_i - \mu_j : i = 1..(n+1) \right\}}{\prod \left\{ \mu_i - \mu_j : i = 1..n, i \neq j \right\}} = \frac{1}{2} \left(\left(\sum_{i=1}^{n+1} \lambda_i - \sum_{i=1}^n \mu_i \right)^2 + \sum_{i=1}^n \mu_i^2 - \sum_{i=1}^{n+1} \lambda_i^2 \right).$$

In particular, deduce that

$$\sum_{j=1}^n \frac{\prod \left\{ 2(i-j) - 1 : i = 1..(n+1) \right\}}{\prod \left\{ 2(i-j) : i = 1..n, i \neq j \right\}} = -\frac{n(n+1)}{2}.$$

10408. Proposed by Peter W. Shor, AT&T Bell Labs, Murray Hill, NJ.

Suppose that a function is defined as follows:

$$\begin{aligned} f(1, 1) &= 1 \\ f(i, 0) &= 0 \quad (i \geq 0) \\ f(1, j) &= 0 \quad (j \geq 2) \end{aligned}$$

and

$$f(i, j) = 3(2i - j - 2)f(i - 1, j) + (i - 2j + 3)f(i - 1, j - 1)$$

otherwise.

(a) Show that

$$\sum_{j=1}^{\lfloor i/2 \rfloor + 1} (-1)^{j+1} f(i, j) = i(i+2)(i+4) \cdots (3i-4).$$

(b) Find a closed form expression for $f(2j - 1, j)$.

Theorem. Let $f(X) = Y$ be light and open where X and Y are 2-manifolds and X is without boundary, let q be any ordinary point of Y and let $p \in f^{-1}(q)$. There exists an integer $k > 0$ such that if B is any sufficiently small closed 2-cell on Y with q in its interior, the component A of $f^{-1}(B)$ containing p is a closed 2-cell mapping onto B by a power map of degree k , i.e., $f|_A$ is topologically equivalent to $w = z^k$ on $|z| \leq 1$ (with p corresponding to 0, i.e., $A \cap f^{-1}(q) = p$).

To explain the terminology here, f is a continuous map of X onto Y , and an ordinary point of Y means one in the interior of a 2-cell contained within Y ; f is light means $f^{-1}(y)$ is totally disconnected for every y in Y ; and f is open means $f(V)$ is open in Y for every open V in X . In the given problem we take $X = U$ (a 2-manifold without boundary) and $Y = f(X)$ and show first that the theorem is applicable. Differentiability of f in the given sense means that the coördinate functions of f have total differentials or equivalently that, for each $x \in U$, $f(x+h) = f(x) + Mh + \eta$ where M is the Jacobian matrix and the term $\eta = o(|h|)$ as $h \rightarrow 0$. Since the Jacobian $J(x) \equiv |M| \neq 0$ it follows that in the neighbourhood of each point of X f behaves, to first order, like a non-singular linear mapping. It follows that for every $x \in U$ there is a neighbourhood N of x such that $x_1 \in N$ and $x_1 \neq x$ implies that $f(x_1) \neq f(x)$, so that for every $y \in Y$, $f^{-1}(y)$ is a set of isolated points and thus totally disconnected. So f is light.

Lemma. The map $f : U \rightarrow \mathbb{R}^2$ is strongly open, that is for each open V in U , $f(V)$ is an open subset in \mathbb{R}^2 . Hence $Y = f(U)$ is open in \mathbb{R}^2 and so is a 2-manifold, and all $f(V)$ are relatively open in Y so that $f : U \rightarrow Y$ is open.

Proof. If f fails to be strongly open there will be a point $x \in U$ and a closed disc N centred on x such that $y = f(x)$ does not lie in the interior of $f(N)$ in \mathbb{R}^2 . N may be supposed small enough to ensure, by the hypotheses on f , that $y \notin f(\partial N)$. Since $f(N)$ and $f(\partial N)$ are closed one may choose $z \in \mathbb{R}^2$ near y so that $0 < d(z, f(N)) < d(z, f(\partial N))$ where d denotes the usual \mathbb{R}^2 metric. If then $d(z, f(N)) = d(z, f(t))$ for some $t \in N$, which exists because $f(N)$ is closed, it follows t lies in the interior of N . But if M is the Jacobian matrix at t , $f(t+h) = f(t) + Mh + \eta$ will certainly be nearer to z than $f(t)$ for all sufficiently small h in a suitable direction, and $t+h$ will be in N , contradicting the definition of $f(t)$ as a nearest point of $f(N)$ to z . So after all f must be strongly open and the lemma is proved.

The theorem is therefore applicable, and since $Y = f(U)$ is open in \mathbb{R}^2 every point of it is ordinary, and every $p \in U$ is contained in the interior of a closed 2-cell A which maps by f onto a closed 2-cell B containing $f(p)$ in its interior, with $f|_A$ topologically equivalent to a power mapping $w = z^k$ on $|z| \leq 1$. Since p corresponds to 0, a small circle centered on p will map to a curve winding k times about q —but since f is a non-singular linear mapping to first order in the neighbourhood of p , the image curve only winds once round q , so that $k = 1$ and $f|_A$ is a homeomorphism of A onto B . If a small circle centre p is imagined oriented anticlockwise then the image curve will be oriented anticlockwise or clockwise according as $J(p) > 0$ or $J(p) < 0$. And similarly for any other point p^* in the interior of A . But if A is a 2-cell then agreement in sense for two simple closed curves interior to A is invariant under a homeomorphism of A (Whyburn, p.63) and it follows that $J(p^*)$ and $J(p)$ have the same sign. Hence $U_+ \equiv \{p : J(p) > 0\}$ and $U_- \equiv \{p : J(p) < 0\}$ are open sets and it follows by connectedness of U that $U = U_+$ or $U = U_-$.

It is of interest to note that the above has actually proved that a differentiable $f : U \rightarrow \mathbb{R}^2$ with non-zero Jacobian is locally a homeomorphism and so one gets a strong version of the inverse function theorem for such mappings, without any continuity assumptions on the partial derivatives of the coördinate functions of f .

There is no record of any other solution being received.

One Norm Doesn't Restrict Algebraic Conjugates

10221 [1992, 461]. *Proposed by Raphael M. Robinson, University of California, Berkeley, CA.*

Let α and β be conjugate algebraic numbers with $|\alpha| = 1$.

(a) Show that if $|\beta| \neq 1$, then $|\beta|^2$ must be irrational.

(b) Show that the possible values of β are everywhere dense in the complex plane.

Solution by Richard Holzsager, The American University, Washington, DC. Since α is on the unit circle, its complex conjugate $1/\alpha$ is among its conjugates over \mathbb{Q} . Therefore each of its conjugates is accompanied by its reciprocal, and so $|\beta|^2 = \beta/\gamma$, where $\gamma = 1/\bar{\beta}$. If this were a rational number r , then γ and $r\gamma$ have the same minimal polynomial $P(x) = x^n + a_{n-1}x^{n-1} + \dots + a_0$, so $a_j = a_j r^{n-j}$, and r (being positive) must be 1.

To see the density of such β , we change the problem a bit. Since the roots occur in reciprocal pairs, $a_j = a_{n-j}$ and n is even. Put $n = 2m$. Then $P(x)/x^m$ is a polynomial $Q(y)$ in $y = x + 1/x$. The polynomial P has a root on the unit circle if and only if Q has a root in $[-2, 2]$. Therefore, if we can show that conjugates b of algebraic numbers a in $[-2, 2]$ are dense in the plane, we can pass to $\beta = (b \pm \sqrt{b^2 - 4})/2$ to get the result for the original situation.

In fact, any interval $(-r, r)$ is sufficient. To see this, consider the conjugate numbers $u = 2^{1/4}$ and $v = ui$. Look at the two-dimensional subspace U of \mathbb{R}^3 consisting of vectors (a, b, c) such that $au + bu^2 + cu^3 = 0$. Since $av + bv^2 + cv^3 = 0$ only for $b = 0$ and $a - 2^{1/2}c = 0$, the mapping $L : U \rightarrow \mathbb{C}$ given by $(a, b, c) \mapsto av + bv^2 + cv^3$ is a one-to-one linear function. Thus L is onto, and varying a, b , and c slightly gives a rational combination of v, v^2 , and v^3 that is arbitrarily close to a given number, while the corresponding combination of u, u^2 , and u^3 is in $(-r, r)$.

Editorial comment. All solvers proved refined versions of (b). Generally, constructions used cubic polynomials in the role of $Q(y)$ to show that sextic irrationals suffice. Robert B. Israel showed that, in addition, α can be restricted to any interval of the unit circle. On the other hand, the proposer and Ilan Kozma showed that, allowing all degrees, α could be required to be an algebraic integer.

Solved also by R. J. Chapman (U. K.), R. B. Israel (Canada), S. Kanetkar, I. Kozma (Israel), O. P. Lossers (The Netherlands), and the proposer.

Ratios in a Matrix Power

10224 [1992, 362]. *Proposed by Yves Nievergelt, Eastern Washington University, Cheney, WA.*

Consider all 2 by 2 real matrices $A = (a_{i,j})$ having non-negative determinant, all entries positive, and $a_{1,1} = a_{2,2}$. Also, for each positive integer p , denote by $a_{i,j}^{(p)}$ the entries of the power A^p . Prove that

$$\lim_{p \rightarrow \infty} \frac{a_{1,1}^{(p)}}{a_{2,1}^{(p)}} = \sqrt{\frac{a_{1,2}}{a_{2,1}}} = \lim_{p \rightarrow \infty} \frac{a_{1,2}^{(p)}}{a_{1,1}^{(p)}}.$$

Solution by Benjamin G. Klein, Davidson College, Davidson, NC. For notational convenience, write $A = \begin{bmatrix} a & b \\ c & a \end{bmatrix}$. We assume only that a and bc are positive. (The assumptions that $a^2 - bc \geq 0$ and $b > 0$ are superfluous.) Let $\lambda = a + \sqrt{bc}$ and $\mu = a - \sqrt{bc}$. Computations involving eigenvalues lead to the following formula, which is readily verified by induction.

$$A^p = \frac{1}{2} \begin{bmatrix} \lambda^p + \mu^p & \sqrt{b/c}(\lambda^p - \mu^p) \\ \sqrt{c/b}(\lambda^p - \mu^p) & \lambda^p + \mu^p \end{bmatrix} \quad \text{for } p = 1, 2, 3, \dots$$

Thus

$$\frac{a_{1,1}^{(p)}}{a_{2,1}^{(p)}} = \sqrt{b/c} \left(\frac{1 + (\mu/\lambda)^p}{1 - (\mu/\lambda)^p} \right) \quad \text{and} \quad \frac{a_{1,2}^{(p)}}{a_{1,1}^{(p)}} = \sqrt{b/c} \left(\frac{1 - (\mu/\lambda)^p}{1 + (\mu/\lambda)^p} \right).$$

Since the absolute value of λ is greater than the absolute value of μ , $(\mu/\lambda)^p \rightarrow 0$ as $p \rightarrow \infty$, and the desired result follows at once.

Solved by 44 readers and the proposer.

A Half-Binomial Identity

10229 [1992, 570]. *Proposed by Herman Bavinck, Delft University of Technology, Delft, The Netherlands.*

Given that m and p are integers with $m \geq p \geq 1$, evaluate

$$\sum_{j=1}^p \binom{1/2}{m-j+1} \binom{1/2}{m+j}.$$

Solution by O. P. Lossers, University of Technology, Eindhoven, The Netherlands. The sum is equal to

$$-\frac{p}{2m(2m+1)} \binom{-1/2}{m-p} \binom{-1/2}{m+p}.$$

This result was found by extensive use of binomial identities, but it is easily verified by induction on p : The result is easily seen to hold for $p = 1$ (or for $p = 0$), and the induction step requires only the straightforward identity

$$\begin{aligned} & \frac{p-1}{2m(2m+1)} \binom{-1/2}{m-p+1} \binom{-1/2}{m+p-1} \\ & - \frac{p}{2m(2m+1)} \binom{-1/2}{m-p} \binom{-1/2}{m+p} = \binom{1/2}{m-p+1} \binom{1/2}{m+p}. \end{aligned}$$

Editorial comment. The identity actually holds for all integers $p \geq 0$. Some more general results were found through the use of identities for hypergeometric series. In particular, François Grondin reduced the formula to a special case of a result of J. L. Lavoie, "Partial sums of coefficients of well-poised hypergeometric series", *Boll. Un. Mat. Ital.* (3) 21 (1966), 346–352. The solution can also be found with Gosper's indefinite summation algorithm, which is implemented in several computer algebra systems.

Solved also by D. Alvis, J. Anglesio (France), R. J. Chapman (U. K.), F. Grondin (Canada), H. van Haeringen (The Netherlands), I. Kastanas, I. Nemes (Austria), D. Zeilberger, and the proposer.

Another Version of the Axiom of Choice

10245 [1992, 675]. *Proposed by M. A. Bezem, Utrecht University, Utrecht, The Netherlands, and A. J. C. Hurkens, Catholic University, Nijmegen, The Netherlands.*

Let \mathcal{S} be a set of finite, non-empty sets. A *transversal* of \mathcal{S} is a set which has a non-empty intersection with every element of \mathcal{S} . The Principle of Minimal Transversal states that every such \mathcal{S} has a transversal that is minimal with respect to set inclusion. Prove that the Axiom of Choice is equivalent to the Principle of Minimal Transversal.

Solution by S. F. Barger, Youngstown State University, Youngstown, OH. For a set \mathcal{S} of finite, non-empty sets, the intersection of a chain of transversal of \mathcal{S} is evidently a transversal of \mathcal{S} , so Zorn's Lemma (equivalent to the Axiom of Choice) implies the Principle of Minimal Transversal. Conversely, for any partially ordered set P , let \mathcal{S} be the set of all 2-element antichains in P . A subset $C \in P$ is a chain if and only if it contains no element of \mathcal{S} ; i.e., if and only if $P - C$ is a transversal of \mathcal{S} . Hence a minimal transversal of \mathcal{S} is the complement of a maximal chain in P , so the Principle of Minimal Transversal implies Zorn's Lemma.

Solved also by R. Holzager, I. Kastanas, O. P. Lossers (The Netherlands), K. Schilling, K. Smith & R. Griffus (students), Western Maryland College Problems group, and the proposers. Four incorrect solutions were received.

Reconstructing an Isosceles Triangle

10249 [1992, 782]. *Proposed by O. Yumlu, Munich, Germany.*

Suppose that the inradius of an isosceles triangle and the ratio of the distances from its incenter to its vertices are given. Give a Euclidean construction of the triangle.

Solution by Paul Yiu, Florida Atlantic University, Boca Raton, FL. Let $\triangle ABC$ be the desired isosceles triangle with $AB = AC$, incenter I and inradius r . Construct an isosceles right triangle $\triangle IXY$ with $IX = XY = r$ and a right angle at X . Let X', X'' be points on the line IX such that $IX' : IX'' = k =$ the given ratio $IA : IB$ (It is assumed that k is given in constructible form). The line through X' parallel to $X''Y$ meets the line IY at a point Y' such that $IY' = \sqrt{2}kr$. On the line XY , mark a point P with $XP = IY' = \sqrt{2}kr$. Let $\mathcal{C}, \mathcal{C}'$ be the circles with center I passing through X and P respectively. Note that a tangent from any point on \mathcal{C}' to \mathcal{C} has square length $2k^2r^2$. Mark a point Q on \mathcal{C} with $XQ = r$, and extend XQ to meet \mathcal{C}' at K . Finally, extend the ray XI to a point A such that $IA = XK$. The triangle bounded by the tangents to \mathcal{C} from A and the tangent at X is the desired isosceles triangle.

To justify the construction, let α and β be the angles at A and B respectively. Since $IA(IA - r) = KX(KX - QX) =$ the square length of the tangents from K to $\mathcal{C} = 2k^2r^2$, we have

$$\frac{r}{\sin \frac{\alpha}{2}} \cdot \left(\frac{r}{\sin \frac{\alpha}{2}} - r \right) = 2k^2r^2, \\ 1 - \sin \frac{\alpha}{2} = 2k^2 \sin^2 \frac{\alpha}{2}. \quad (1)$$

Since $\frac{\alpha}{2} + \beta = \frac{\pi}{2}$,

$$\sin \frac{\alpha}{2} = \cos \beta = 1 - 2 \sin^2 \frac{\beta}{2}. \quad (2)$$

From (1) and (2) it follows that $\sin \frac{\beta}{2} = k \cdot \sin \frac{\alpha}{2}$, and $IA = k \cdot IB$.

Solved also by E. Alkan (student, Turkey), N. K. Artemiadis (Greece), S. F. Barger, A. Berenstein (student, Colombia), M. V. Bjelica (Yugoslavia), R. J. Chapman (U. K.), A. Coffman, I. Dimitrić, J. Dou (Spain), J. Fukuta (Japan), H. Helfgott (student, Peru), I. Kastanas, O. P. Lossers (The Netherlands), J. McHugh, P. Ranaldi, G. Velissarios (Greece), and the Anchorage Math Solutions Group. One incorrect solution was received.

Collaborating editors: David F. Appleyard, Paul T. Bateman, Bruce C. Berndt, Duane M. Broline, Barry W. Brunson, Frank S. Cater, Gulbank D. Chakerian, Underwood Dudley, Gerald A. Edgar, Michael A. Filaseta, Ira M. Gessel, Richard A. Gibbs, Jerrold R. Griggs, Douglas A. Hensley, John R. Isbell, Mourad E. H. Ismail, Murray Klamkin, Daniel J. Kleitman, Frederick W. Luttman, Frank B. Miles, Richard Plieffer, Stephen L. Portnoy, J. O. Shallit, John Henry Steelman, Kenneth B. Stolarsky, David E. Tepper, Douglas B. Tyler, Daniel Ullman, and William E. Watkins.

REVIEWS

Edited by **Darrell Haile**
Indiana University, Bloomington IN 47405

Brouwer's Intuitionism. By Walter P. van Stigt, Elsevier Science, New York, 1990, xxvi + 532 pp, \$94.70.

Reviewed by **C. Smoryński**

Some years ago my topology teacher told the class that Brouwer had been a competent topologist who went off the deep end and invented intuitionism. More recently, in *The Emperor's New Mind*, Roger Penrose retold the same story in greater detail, if a bit more politely, with his remark that "intuitionism was initiated, in 1924, by the Dutch mathematician L. E. J. Brouwer as an alternative response—distinct from that of formalism—to the paradoxes (such as Russell's) that can arise when too free a use of infinite set is made in mathematical reasoning." Concerning the firing of Brouwer from the editorial board of *Mathematische Annalen*, Constance Reid, in her biography of David Hilbert, tells us that this was because Brouwer insisted on handling all papers on topology or of Dutch authorship and that he caused unconscionable delays with these papers. Simultaneously, in Abraham Fraenkel's autobiography, we can learn that Brouwer was fired because he rejected too many papers written by Russian-Jewish mathematicians. Fraenkel does, however, compare Brouwer favourably with van der Waerden, who remained in Nazi Germany throughout the war. While van der Waerden has been reconciled with his compatriots, Brouwer is still widely believed to have been a Nazi, and indeed the pre-publication version of one major obituary of him reported that Brouwer had joined the SA and marched in its parades, his hair flowing majestically in the wind.

Brouwer was not a topologist before becoming an intuitionist. He did not become an intuitionist in response to the paradoxes. His firing from the editorial board of *Mathematische Annalen* had nothing to do with either his performance or any antisemitism on his part. And, there is no evidence whatsoever that Brouwer was ever a Nazi. There are kernels of truth in some of these stories, but no real fruit. The transition from topology to intuitionism is only apparent. Brouwer's dissertation of 1907 had been on the foundations of mathematics. In 1908 he published two papers on intuitionistic lines, one rejecting the higher infinite cardinals and one questioning the validity of the Law of Excluded Middle in infinite domains. Thereafter, at the suggestion of his advisor, D. J. Korteweg (of the Korteweg-de Vries equation), Brouwer worked in classical mathematics, single-handedly developing algebraic topology. Brouwer seemingly abandoned intuitionism for several years for the sake of job procurement and security. This seeming desertion was, however, merely apparent. All along, Brouwer developed the type of topology which he believed could be rendered intuitionistically acceptable. After he had become an internationally recognised mathematician, he began his return to intuitionism, his historically most important paper being published in

two parts in 1918 and 1919 (not 1924) and accompanied by commentaries in Dutch and German journals. In these papers he rejected outright the application of the Law of the Excluded Middle in infinite domains, and yet was greatly successful in developing set theory intuitionistically.

The true reasons behind Brouwer's firing from the board of editors of *Mathematische Annalen* have already been discussed by Dirk van Dalen in the *Mathematical Intelligencer* (vol. 12, #4 (1990), pp. 17–31) and need not be repeated here. Suffice it to say that Constantin Caratheodory, one of the senior editors, considered Brouwer to be the most conscientious editor on the staff, and that, when Brouwer subsequently founded his own journal, his Nazi friend Ludwig Bieberbach was displeased by the number of Jewish editors the journal had. It is true that Brouwer had Nazi friends and that, because of property owned in Germany, he was slow to divorce himself from the regime. It is also true that he refused to break off his friendship with one Dutch mathematician who had been a Nazi during the war and even tried to find the fellow a position afterwards. But none of this makes Brouwer himself a Nazi, and he was even cleared by the Dutch government after the war.

The sources of some of the misinformation on Brouwer are obvious. Brouwer's influential paper on intuitionism—the one that converted Hermann Weyl and precipitated the notorious *Grundlagenstreit* of the 1920s—came after his famous work on algebraic topology. The conclusion drawn by my topology teacher and Penrose is an immediate corollary. That Brouwer's intuitionism was a response to the paradoxes is a corollary to the general Theorem that all foundational work is a response to Russell's paradox. Even Hilbert, who discovered paradoxes in the mid-to-late 1890s, whose student Ernst Zermelo had discovered Russell's paradox by the late 1890s, who publicly cited a consequence of the paradoxes in his address to the International Congress of Mathematicians in 1900, and who himself was not unduly moved by them (he didn't even bother to mention them in 1900) until he saw Russell's and Dedekind's overreactions to them—even Hilbert, I say, who knew that Leopold Kronecker died in 1891 before the paradoxes were known, concluded that Kronecker's rejection of set theory was a response to the paradoxes. In point of fact, the paradoxes had little direct impact. Russell, Frege, Dedekind, and Weyl took them seriously. Hilbert took the reaction to them seriously, but was not directly moved by them. He only sprang into action when Weyl was converted to intuitionism, in which approach to mathematics the paradoxes were irrelevant. Hilbert's foundational work of the 1920s was a response to intuitionism, not to the paradoxes. Yet he paraded it as a response to the paradoxes, expanding their role to include Kronecker and, by implication, Brouwer.

The fact of the matter is that much of what we know about Brouwer was written by Hilbert, a great mathematician, but an unreliable historian. If, for example, we read that Brouwer attacked Hilbert's formalism as a “formula game,” we can trace the accusation to one of Hilbert's papers. We will not find it in any of Brouwer's papers, but we will find that Hermann Weyl used the phrase! In all of Hilbert's papers of the 1920s there is, in fact, but a single direct reference to any of Brouwer's writings—and this concerns the possibility of finding decimal expansions of real numbers. Every other reference to Brouwer by Hilbert is either a reference to something Weyl said or Kronecker did; it is not a direct reference to Brouwer.

If one reads the papers of Hilbert and Brouwer of the 1920s, one begins to doubt there ever was a *Grundlagenstreit* at all. One finds Hilbert bravely fighting the ghost of Kronecker, or a cardboard replica of Brouwer unwittingly fabricated

by Weyl. Those familiar with the anecdotes about Kronecker or the more recent antics of Bishop will be surprised to find that Brouwer's papers of the period are not the least bit dogmatic. Unlike Kronecker who argued *ad hominem*, or Weyl and Bishop who simply labelled absurd that with which they disagreed, Brouwer offered careful criticism of formalism and argued dispassionately for intuitionism. The only show of temper on Brouwer's part in the foundational struggle occurred late, when Brouwer declared that formalism need not sneer at intuitionism considering both that it had not accomplished anything (true enough) and that it had plagiarised intuitionism (very probably true, but not yet proven to the historian's satisfaction). Perhaps what Brouwer said in public differs from what he wrote; if so, I would like to learn of it. For the time being, however, it seems to me that Brouwer's reputation for the sort of behaviour correctly associated with constructivist-fascists like Kronecker and Bishop rests entirely on Hilbert's accusations based on Weyl's bombastic misrepresentations.

I hesitate to say that it is time to rehabilitate Brouwer. He was, after all, no saint. He did battle with Hilbert (albeit only on tactics to better the treatment of German scientists after the First World War). He fought numerous battles over priority and could be quite petty in such fights. He was also a rabid sexist. We certainly have no cause to revere the man; but we should not revile him either. Brouwer, the founder of algebraic topology and of one of the few coherent philosophies of mathematics, was one of the great mathematicians of our century, the greatest Holland has produced since Christiaan Huyghens, and he deserves better from the mathematical community than he has received.

With some reservations I would refer the curious or even repentant reader to van Stigt's book for a more accurate picture of Brouwer than his topology teacher may have given him or he may have presented to his own students. The most serious reservation is that the book is a work on the philosophy of mathematics, not the history thereof, and is not, therefore, the appropriate choice for the mathematical community. It is to a large extent, however, the only choice currently available and we must consider it. Only the first chapter, semi-biographical in character, is of direct relevance to the mathematician. This biography, however sketchy, is the most extensive one available in English and, however much it focuses on philosophy, it does touch on Brouwer's mathematical work, particularly on his topology, and its relation to his philosophy. The chapter certainly ought to correct the widespread misunderstanding about the late origin of Brouwer's foundational interests or their being responses to the paradoxes. The discussion of the *Grundlagenstreit*, of necessity short, will do nothing to replace the popular fiction and must therefore be considered inadequate for revising popular Brouwerian history. (Proof of this inadequacy is at hand: Barrow's recent *Pi in the Sky* relies heavily on the present book and treats Brouwer as not only an active participant in the fight, but also as the villain.) Van Stigt does run rings around Reid and Fraenkel on the matter of Brouwer's firing from *Mathematische Annalen*, but van Dalen's already cited *Intelligencer* article is superior. As to Brouwer's postwar difficulties, van Stigt has the right attitude, but offers no details. All of this, of course, is not a criticism of the book, but indication of how it doesn't serve a purpose for which it wasn't intended—whence the reservation cited above.

Brouwer's Intuitionism itself cannot, however, escape criticism. Reviewers have made it clear that it is not entirely reliable. It was clearly not proofread and may not even have been edited. The fact, for instance, that repeated Brouwerian quotations are translated anew on repetition suggests this latter. Conclusions and speculations are offered as facts on inadequate grounds. For example, in discussing

the *Mathematische Annalen* affair, van Stigt quotes Brouwer's speculation that Hilbert's motive was anger at Brouwer's declining a position in Göttingen in 1919 and, instead of critically examining this or recalling other documented sources of Hilbert's anger (which would suggest Brouwer was not completely right in his estimation), van Stigt supports the assertion with further speculation. As an indication of just how far from the mark such speculation can get, I might mention van Stigt's remark on Brouwer's fellowship with various young poets: "Brouwer lacked the ability to express his feelings artistically". This certainly runs counter to what every Dutchman I've spoken to has said of Brouwer's correspondence with Adama van Scheltema: Brouwer's prose is uniformly favourably compared with that of the poet. One criticism I've read and must agree with is that van Stigt's psychological analysis of Brouwer is excessive and excessively negative. Indeed, it might be said that van Stigt has done a hatchet job on Brouwer. But, in doing so, he has still managed to paint a more positive image of Brouwer than is generally displayed to mathematicians and the reader can get a first, albeit rough, approximation to Brouwer from van Stigt's book. Thus, with the reservations that the book is only tangentially relevant to the mathematical community and that it must be approached very critically, I refer the reader to it for a faltering first step in getting to know the real Brouwer.

429 South Warwick Avenue
Westmont, IL 60559

There are in this world optimists who feel that any symbol that starts off with an integral sign must necessarily denote something that will have every property that they should like an integral to possess. This of course is quite annoying to us rigorous mathematicians; what is even more annoying is that by doing so they often come up with the right answer.

—E.J. McShane

Bulletin of the American Mathematical Society
v 69, p 611, 1963.

**Answer to Picture Puzzle
(p. 770)**

No, they are not: they are Emile Borel and Armand Borel.

TELEGRAPHIC REVIEWS

Edited by **Arnold Ostebee and Paul Zorn**

with the assistance of the Mathematics Departments of
Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

<i>T</i> : Textbook	<i>P</i> : Professional Reading	1-4: Semester
<i>C</i> : Computer Software	<i>L</i> : Undergraduate Library	** : Special Emphasis
<i>S</i> : Supplementary Reading	13: Grade Level	?? : Questionable

Readers are advised that price information is subject to change. Selected books and software packages receive a second, more extensive review in the *Monthly*.

Books and software submitted for review should be sent to *Book Reviews Editor*, *American Mathematical Monthly*, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098.

General, P, L. *Comic Sections: The Book of Mathematical Jokes, Humour, Wit and Wisdom.* Desmond MacHale. Boole Pr, 1993, v + 152 pp, (P). [ISBN 1-85748-007-4; 1-85748-006-6] A considerable compendium of jokes, quotations, cartoons, excerpts, bloopers, verse, and historical anecdotes on (loosely) mathematical themes. Inevitably, the amusement index varies wildly, but on balance a valuable resource for common rooms, department chairs, and the conversationally challenged. **PZ**

General, P, L.** *Proofs Without Words: Exercises in Visual Thinking.* Roger B. Nelsen. Classroom Resource Materials, No. 1. MAA, 1993, ix + 152 pp, \$23 (P). [ISBN 0-88385-700-6] A delightful collection of "pictures or diagrams that help the observer see *why* a particular statement may be true, and also to see *how* one might go about proving it true." Sections on geometry and algebra; trigonometry, calculus and analytic geometry; inequalities; integer sums; sequences and series. **AO**

General, L*. *The Visual Mind: Art and Mathematics.* Ed: Michele Emmer. MIT Pr, 1993, xvii + 274 pp, \$39.95. [ISBN 0-262-05048-X] 35 essays by mathematicians and visual artists, in 4 sections: geometry and visualization; computer graphics, geometry and art; symmetry; perspective, mathematics, and art. Many color illustrations. A fascinating approach to both the mathematics of visual art and the visualization of mathematics, and a valuable resource for mathematics appreciation courses. **AO**

Reference, P, L*. *L^AT_EX: A Document Preparation System User's Guide and Reference Man-*

ual, Second Edition. Leslie Lamport. Addison-Wesley, 1994, xvi + 272 pp, (P). [ISBN 0-201-52983-1] Reflects changes introduced with L^AT_EX 2_ε, the new standard version of L^AT_EX. (*First Edition*, TR, November 1986.) **AO**

Reference, P*, L*. *The L^AT_EX Companion.* Michel Goossens, Frank Mittelbach, Alexander Samarin. Addison-Wesley, 1994, xxx + 528 pp, (P). [ISBN 0-201-54199-8] Advanced guide to L^AT_EX 2_ε and to 150+ packages that extend or modify basic L^AT_EX system (e.g., the New Font Selection Scheme, *AMS-L^AT_EX*). Essential for serious L^AT_EX users. **AO**

Reference, C, P. *Math Into T_EX: A Simple Introduction to AMS-L^AT_EX.* George Grätzer. Birkhäuser, 1993, xxix + 294 pp, \$42.50 (P), disk included. [ISBN 0-8176-3637-4] Describes and illustrates *AMS-L^AT_EX* (a high-level mathematical document processor, built atop T_EX) for both rank beginner and seasoned T_EX and L^AT_EX users. Three parts: basic introduction (including installation on PC, Macintosh); comprehensive reference manual; advanced topics. Complete index, useful appendices. Informal exposition; many examples. Many sample files provided on diskette. **PZ**

Reference, P, L. *Memorabilia Mathematica: The Philomath's Quotation Book.* Robert Edouard Moritz. Spectrum Ser. MAA, xiii + 410 pp, \$24 (P). [ISBN 0-88385-513-5] Reprint of 1914 compilation of 1140 Anecdotes, Aphorisms, and Passages by Famous Mathematicians, Scientists & Writers. Companion to *Out of the Mouths of Mathematicians*. **BC**

Reference, P. L. *Out of the Mouths of Mathematicians: A Quotation Book for Philomaths.* Rosemary Schmalz. Spectrum Ser. MAA, 1993, x + 294 pp, \$29 (P). [ISBN 0-88385-509-7] Companion to *Memorabilia Mathematica*, with up-to-date quotes. This reviewer's favorite, due to Mark Kac: "There are worse things than being wrong, and being dull and pedantic are surely among them." BC

Recreational Mathematics, P. *Challenging Puzzles.* Colin Vout, Gordon Gray. Cambridge Univ Pr, 1993, 126 pp, \$14.95 (P). [ISBN 0-521-44602-3] 100 interesting and amusing puzzles, with hints and solutions. DH

Recreational Mathematics, P. *How Amazing.* Charles Snape, Heather Scott. Cambridge Univ Pr, 1992, 48 pp, \$9.95 (P). [ISBN 0-521-35672-5] Colorful photographs and drawings of famous mazes and related puzzles. Elementary combinatorial situations, topological tricks, and network problems. Fun for everyone; useful resource for middle school teachers. MW

Elementary, T*(13: 1). *Intermediate Algebra: Graphs and Functions.* Roland E. Larson, et al. DC Heath, 1994, xxii + 905 pp. [ISBN 0-669-33755-2] Emphasizes problem solving skills, algebraic modeling. Includes discussion problems, optional sections on graphing calculators, computer graphing. TH

Elementary, S(15-18), C. *NUMERATOR in the Mathematics Classroom.* Peter Armstrong, et al. Comp. in Math. Teach. Ser. Mathematical Assoc, 1993, iv + 20 pp, \$3.75 (P). [ISBN 0-906588-30-8] Report on pros and cons of NUMERATOR, British software that simulates function machines. Makes comparisons with spreadsheets, LOGO, BASIC. MW

Education, S(15-18). *Number Sense Now! Reaching the NCTM Standards.* Proj. Dir.: Francis Fennell. NCTM, 1993, xiv + 128 pp, \$98 (P), with 3 videos. [ISBN 0-87353-361-5] Three half-hour classroom vignettes, showing: (1) nature and importance of number sense; (2) instructional strategies, emphasizing communication and hands-on activities; (3) real-life applications. *Guidebook* contains vignette summaries, background information, and blackline masters. Useful for pre-service and in-service elementary teachers. MW

Education, S(17-18), P. *Rethinking Elementary School Mathematics: Insights and Issues.* Eds: Terry Wood, et al. J. for Res. in Math. Educ., No. 6. NCTM, 1993, vi + 122 pp, \$7 (P). [ISBN 0-87353-362-3] Report on a field-based study of second-graders' mathematical learning in a constructivist-oriented classroom. Data collection from whole-class

teaching experiments, individual clinical interviews, and pair collaborations. Methodology may serve as a model for other studies. Conclusions support blending cognitive and sociological interpretations of mathematical activity. Final chapters offer advice for educational innovators and researchers. MW

Education, S(15-17), C. *Spreadsheets: Exploring their Potential in Secondary Mathematics.* David Green, Peter Armstrong, Richard Bridges. Comp. in Math. Teach. Ser. Mathematical Assoc, 1993, ix + 154 pp, \$8.50 (P). [ISBN 0-906588-29-4] Basic concepts and example spreadsheets for arithmetic, algebra, statistics. Ideas for dynamic modeling and simulation. Disk available for Excel. MW

Education, S(15-17). *Graphic Calculators in the Mathematics Classroom.* Peter Arter, et al. Comp. in Math. Teach. Ser. Mathematical Assoc, 1993, vii + 124 pp, \$6.75 (P). [ISBN 0-906588-31-6] Introduction to use of calculators; keystroke sequences and programs for Casio fx-7000, fx-7700, and TI-81. Ideas for exploring algebra, graphing, and statistics. Organization sometimes confusing. MW

Education, S(15-17), P. *Reforming Science Education: Social Perspectives and Personal Reflections.* Rodger W. Bybee. Ways of Knowing in Sci. Ser. Teachers College Pr, 1993, xvii + 197 pp, \$19.95 (P). [ISBN 0-8077-3260-5] 10 provocative essays, written over 15 years, on precollege science education reform. AO

Education, S(14-18). *Activities for Active Learning and Teaching: Selections from the Mathematics Teacher.* Eds: Christian R. Hirsch, Robert A. Laing. NCTM, 1993, v + 244 pp, \$11 (P). [ISBN 0-87353-363-1] 35 activities on problem solving, numeracy, algebra and graphs, geometry and visualization, data analysis and probability. Some computer activities updated to use graphing calculators. Resource for teachers of grades 7-10. MW

Education, S. *Self-Directed Problem Solving: Idea Production in Mathematics.* M. Ann Dirkes. University Pr of America, 1993, 123 pp, \$17.50 (P). [ISBN 0-8191-9130-2] Describes a program in which "students aggressively pursue what they already know and use it to construct new concepts." Primary vehicle seems to be use of various brainstorming techniques. Claims to be effective in secondary school and college mathematics courses, but mathematical content is arithmetic. MW

Education, S(17-18), P. *Reconstructing Mathematics Education: Stories of Teachers Meeting the Challenge of Reform.* Deborah Schifter, Catherine Twomey Fosnot. Teachers College

Pr, 1993, xv + 215 pp, \$18.95 (P). [ISBN 0-8077-3206-0] Stories of teachers implementing constructivist-based teaching strategies in elementary classrooms. Valuable resource for in-service programs. MW

Education, P. *Classroom Dynamics: Implementing a Technology-Based Learning Environment*. Ellen B. Mandinach, Hugh F. Cline. Lawrence Erlbaum Assoc, 1994, xvi + 211 pp, \$49.95. [ISBN 0-8058-0555-9] Case study of a project using simulation software in high school science. Reveals curriculum and teaching issues inherent in use of technology. MW

History, P, L.** *Golden Years of Moscow Mathematics*. Eds: Smilka Zdravkovska, Peter L. Duren. History of Math., V. 6. AMS, 1993, ix + 271 pp, \$94. [ISBN 0-8218-9003-4] 11 essays/reminiscences on mathematical life at Moscow State University in the half-century following the Revolution. Fascinating glimpse of a vital and fertile mathematical milieu. SK

History, S(13–17), L**.** *Ars Magna or The Rules of Algebra*. Girolamo Cardano. Transl: T. Richard Witmer. Dover, 1993, xxiv + 267 pp, \$8.95 (P). [ISBN 0-486-67811-3] Translation of classic Renaissance text from 1545, using modern algebraic notation. A must for anyone interested in mathematical history. (1968 MIT edition, TR, June–July 1969.) DP

History, S, P, L*. *Lectures in the History of Mathematics*. Henk J.M. Bos. Hist. of Math., V. 7. AMS, 1993, x + 197 pp, \$86. [ISBN 0-8218-9001-8] 11 thought-provoking, insightful, informative lectures on topics from 17th and 18th century mathematics. BH

Combinatorics, T(17–18: 1), S, P, L. *Complexity: Knots, Colourings and Counting*. D.J.A. Welsh. London Math. Soc. Lect. Note Ser., V. 186. Cambridge Univ Pr, 1993, viii + 163 pp, \$37.95 (P). [ISBN 0-521-45740-8] Readable treatment of algorithmic aspects of knot theory, combinatorics, statistical physics. A clear and up-to-date survey of what's known—and what's still unknown. BC

Combinatorics, T(17: 1), P. *The Theory of Finite Linear Spaces: Combinatorics of Points and Lines*. Lynn Margaret Batten, Albrecht Beutelspacher. Cambridge Univ Pr, 1993, x + 214 pp, \$49.95. [ISBN 0-521-33317-2] Covers work on combinatorial problems in linear spaces including embedding higher dimensional linear spaces in projective spaces. Presents recent results using group theory. With exercises, open problems. DH

Combinatorics, P. *Finite and Infinite Combinatorics in Sets and Logic*. Eds: N.W. Sauer, R.E. Woodrow, B. Sands. NATO ASI Ser. C,

V. 411. Kluwer Academic, 1993, xvii + 453 pp, \$214. [ISBN 0-7923-2422-6] Proceedings of a 1991 NATO Advanced Study Institute in Banff, Canada.

Number Theory, S(15–16), L. *Solved and Unsolved Problems in Number Theory*. Daniel Shanks. Chelsea, 1993, xiv + 305 pp, \$27.50. [ISBN 0-8284-2297-X] The *Fourth Edition* of this unique and engaging book (*Third Edition*, TR, June–July 1986). Only one new page in this edition, but still a bargain. SG

Linear Algebra, S(14–16: 1). *Linear Algebra with DERIVE*. Benny Evans, Jerry Johnson. Wiley, 1994, x + 280 pp, \$20.95 (P). [ISBN 0-471-59194-7] Exercises, solutions, and introduction to DERIVE. Exercises include standard problems, problems not solvable by hand, and exploratory applications. DS

Linear Algebra, T*(13: 1). *Elementary Linear Algebra, Seventh Edition*. Howard Anton. Wiley, 1994, xvi + 620 pp, \$67.95. [ISBN 0-471-58742-7] Now with early introduction to linear transformations, eigenvalues, and characteristic equations in \mathbb{R}^n , more visualization in \mathbb{R}^2 and \mathbb{R}^3 , more proofs tailored for beginners. (*Sixth Edition*, TR, October 1991.) TH

Linear Algebra, T(13–14: 1, 2). *Linear Algebra with Applications, Third Edition*. W. Keith Nicholson. PWS, 1995, xviii + 540 pp. [ISBN 0-534-93666-0] "This edition continues the trend toward spending more time on matrix computations as well as applications, a view supported by the Linear Algebra Curriculum Study Group." Inner product spaces now follow diagonalization, linear transformations. No computer exercises. (*Second Edition*, TR, May 1990.) HD

Group Theory, T*(18: 1), P. *Local Representation Theory*. J.L. Alperin. Stud. in Adv. Math., V. 11. Cambridge Univ Pr, 1993, x + 178 pp, \$24.95 (P). [ISBN 0-521-44926-X] Paperback version of 1986 text (TR, April 1987). DP

Group Theory, P. *Linear Algebraic Groups and Their Representations*. Eds: Richard S. Elman, Murray M. Schacher, V.S. Varadarajan. Contemp. Math., V. 153. AMS, 1993, xii + 200 pp, \$42 (P). [ISBN 0-8218-5161-6] Proceedings of a 1992 conference at UCLA.

Algebra, P. *Symmetries in Science VI: From the Rotation Group to Quantum Algebras*. Ed: Bruno Gruber. Plenum Pr, 1993, xvii + 770 pp, \$149.50. [ISBN 0-306-44584-0] Proceedings of a 1992 symposium in Bregenz, Austria.

Algebra, T(15–16: 1, 2). *Contemporary Abstract Algebra, Third Edition*. Joseph A. Gallian. DC Heath, 1994, xix + 577 pp. [ISBN 0-

669-33907-5] Changes from earlier editions include writing functions to the left of their arguments, chapter on symmetry and counting, new exercises, some internal reorganization. (First Edition, TR, May 1986.) JS

Algebra, T(18: 1, 2), P. *Cohen-Macaulay Rings*. Winfried Bruns, Jürgen Herzog. Stud. in Adv. Math., V. 39. Cambridge Univ Pr, 1993, xi + 403 pp, \$79.95. [ISBN 0-521-41068-1] Self-contained introduction to homological and combinatorial aspects of commutative algebra: Goldstein rings, local cohomology, canonical modules, Hilbert functions, Stanley-Reisner rings, semi-group rings, and determinantal rings. TH

Algebra, T(18), P*. *Polynomial Invariants of Finite Groups*. D.J. Benson. London Math. Soc. Lect. Note Ser., V. 190. Cambridge Univ Pr, 1993, ix + 118 pp, \$32.95 (P). [ISBN 0-521-45886-2] Studies the action of a finite group on the ring of polynomial functions on a linear representation, both in characteristic zero and p . Restriction to finite groups makes the material accessible to graduate students. DP

Calculus, S(13-14). *HP-48G/GX Calculator Enhancement for Science and Engineering Mathematics*. Ed: Donald R. LaTorre. Saunders College, 1994, xii + 418 pp, \$12 (P). [ISBN 0-03-097000-8] 6 chapters on uses of HP-48G/GX calculators in single-variable calculus, multivariable calculus, differential equations, linear algebra, engineering mathematics, probability and statistics. Examples, exercises, activities, explorations, and projects. HD

Complex Analysis, S(18), P. *An Introduction to Sato's Hyperfunctions*. Mitsuo Morimoto. Transl: Mitsuo Morimoto. Transl. of Math. Mono., V. 129. AMS, 1993, xii + 273 pp, \$87. [ISBN 0-8218-4571-3] English translation of 1976 Japanese original. Assumes substantial background in complex variables, topological linear spaces, functional analysis. Treats theory of hyperfunctions of one and several variables, cohomology groups, microfunctions. JS

Dynamical Systems, T(17), P. *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics, Second Edition*. Andrzej Lasota, Michael C. Mackey. Appl. Math. Sci., V. 97. Springer-Verlag, 1994, xiv + 472 pp, \$49. [ISBN 0-387-94049-9] Originally published as *Probabilistic Properties of Deterministic Systems* (TR, November 1987). New material includes chapter on evolution of distributions. SK

Dynamical Systems, P. *Bifurcations and Periodic Orbits of Vector Fields*. Ed: Dana Schlomiuk. NATO ASI Ser. C, V. 408. Kluwer Academic, 1993, xvii + 472 pp, \$180. [ISBN

0-7923-2392-0] 10 papers from the 1992 NATO Advanced Study Institute at Université Montréal.

Dynamical Systems, T?, S(16-18), P, L. *Discrete Iterated Function Systems*. Mario Peruggia. AK Peters, 1993, xvi + 180 pp, \$34.95. [ISBN 1-56881-015-6] Iterated Function Systems (IFSs) can be used to encode digitized color images via a method popularized in Barnsley's *Fractals Everywhere*. Book compares continuous and discrete IFSs; explores ramifications of using one to approximate the other. Background needed: topology and metric spaces, probability, Markov chains. KS

Numerical Analysis, P. *Box Splines*. C. de Boor, K. Höllig, S. Riemenschneider. Appl. Math. Sci., V. 98. Springer-Verlag, 1993, xvii + 200 pp, \$34. [ISBN 0-387-94101-0] Theory of box splines (compactly-supported, smooth, piecewise polynomial functions). DH

Numerical Analysis, T(15: 1). *Numerical Mathematics and Computing, Third Edition*. Ward Cheney, David Kincaid. Brooks/Cole, 1994, xiv + 578 pp, \$60.75. [ISBN 0-534-20112-1] Changes include pseudocode algorithms, examples solved using either Maple or MATLAB. (First Edition, TR, April 1982.) DH

Numerical Analysis, P. *Exploiting Symmetry in Applied and Numerical Analysis*. Eds: Eugene L. Allgower, Kurt Georg, Rick Miranda. Lect. in Appl. Math., V. 29. AMS, 1993, xii + 457 pp, \$58 (P). [ISBN 0-8218-1134-7] Proceedings of a 1992 AMS-SIAM Summer Seminar at Colorado State University.

Functional Analysis, T(18: 1), P. *Introduction to Regularity Theory for Nonlinear Elliptic Systems*. Mariano Giaquinta. Lect. in Math. Birkhäuser, 1993, viii + 131 pp, \$29 (P). [ISBN 0-8176-2879-7] Lecture notes from a winter 1983 course at ETH, Zurich. Requires strong background in graduate real analysis. Works through Hilbert-space regularity, Schauder-estimates and L^p theory to regularity in both scalar- and vector-valued cases. SM

Analysis, P. *Different Perspectives on Wavelets*. Ed: Ingrid Daubechies. Proc. of Symp. in Appl. Math., V. 47. AMS, 1993, xi + 205 pp, \$45. [ISBN 0-8218-5503-4] 8 lectures from a 1993 AMS short course in San Antonio. BH

Algebraic Geometry, T(17-18: 1). *Algebraic Varieties*. George R. Kempf. London Math. Soc. Lect. Note Ser., V. 172. Cambridge Univ Pr, 1993, x + 163 pp, \$37.95 (P). [ISBN 0-521-42613-8] Introduces theory of algebraic functions on varieties from sheaf-theoretic standpoint. Sheaves and sheaf cohomology are presented from the beginning; curves are used to

illustrate the theory. Excellent background for study of schemes. DS

Algebraic Geometry, P. *Moduli Spaces of Abelian Surfaces: Compactification, Degenerations, and Theta Functions.* Klaus Hulek, Constantin Kahn, Steven H. Weintraub. Expos. in Math., V. 12. Walter de Gruyter, 1993, xii + 347 pp, DM 168. [ISBN 3-11-013851-4] Main themes: constructing toroidal compactifications of moduli spaces; constructing degenerate abelian surfaces using ideas of Mumford; geometry of Horrocks-Mumford bundle. RM

Differential Geometry, P. *The Penrose Transform and Analytic Cohomology in Representation Theory.* Eds: Michael Eastwood, Joseph Wolf, Roger Zierau. Contemp. Math., V. 154. AMS, 1993, x + 259 pp, \$47 (P). [ISBN 0-8218-5176-4] 15 papers from a 1992 AMS-IMS-SIAM Summer Research Conference at Mt. Holyoke College.

Differential Geometry, T(18), S, P. *Global Affine Differential Geometry of Hypersurfaces.* An-Min Li, Udo Simon, Guosong Zhao. Expos. in Math., V. 11. Walter de Gruyter, 1993, xiii + 328 pp, DM 178. [ISBN 3-11-012769-5] Equiaffine hypersurface theory, from graduate level to latest results. First local theory, then global results: affine hyperspheres, rigidity and uniqueness theorems, variational problems and affine maximal surfaces, geometric inequalities. No exercises; large bibliography. SG

Geometry, P. *Finite Geometry and Combinatorics: The Second International Conference at Deinze.* Eds: A. Beutelspacher, et al. London Math. Soc. Lect. Note Ser., V. 191. Cambridge Univ Pr, 1993, ix + 412 pp, \$39.95 (P). [ISBN 0-521-44850-6] 35 papers from a 1992 conference in Astene-Deinze, Belgium.

Algebraic Topology, T(17-18: 1). *Fibre Bundles, Third Edition.* Dale Husemoller. Grad. Texts in Math., V. 20. Springer-Verlag, 1994, xix + 353 pp, \$54. [ISBN 0-387-94087-1] Complete guide to theory of fibre bundles, from homotopy theory through characteristic classes. Essential reference for K-theory, classical groups, differentiable manifolds, and many other subjects from algebraic topology. (Second Edition, TR, January 1976.) DS

Topology, P. *Topics in Knot Theory.* Ed: M.E. Bozhüyük. NATO ASI Ser. C, V. 399. Kluwer Academic, 1993, xiv + 353 pp, \$149. [ISBN 0-7923-2285-1] 14 papers from the September 1992 NATO Advanced Study Institute in Erzurum, Turkey.

Optimization, T(16-17). *Interior-Point Polynomial Algorithms in Convex Programming.* Yurii Nesterov, Arkadii Nemirovskii. Stud. in

Appl. Math., V. 13. SIAM, 1994, ix + 405 pp, \$68.50. [ISBN 0-89871-319-6] Genuinely new in setting a body of material (optimization, stressing non-linear convex problems) in a unified context. Suggests new methods for many traditional problems. AWR

Stochastic Processes, T(17), S. *Lecture Notes in Control and Information Sciences-194: Realization Probabilities: The Dynamics of Queuing Systems.* Xi-Ren Cao. Springer-Verlag, 1994, x + 320 pp, \$78 (P). [ISBN 0-387-19872-5] Dynamic approach to steady-state sensitivity formulas. RWJ

Elementary Statistics, S*(7-12). *From Home Runs to Housing Costs: Data Resource for Teaching Statistics.* Ed: Gail Burrill. Dale Seymour Pub, 1994, 110 pp, (P). [ISBN 0-86651-734-0] Interesting data sets from government, media, and other sources, designed for use with Dale Seymour's Quantitative Literacy Series. With suggestions for data analysis, simulation activities, lessons, student projects. RSK

Elementary Statistics, T(13), L. *Exploring Measurements.* Peter Barbella, James Kepner, Richard Scheaffer. Dale Seymour Pub, 1994, vii + 120 pp, (P). [ISBN 0-86651-639-5] Part of Quantitative Literacy Series developed by NCTM, American Statistical Association for school use. Nicely covers measures of center and variability, random sampling, mean and standard deviation of sample average, confidence intervals for population means; uses various small data sets (e.g., sports, weather). RWJ

Elementary Statistics, S(13-16). *Understanding Probability and Statistics: A Book of Problems.* Ruma Falk. AK Peters, 1993, xiii + 239 pp, \$39.95. [ISBN 1-56881-018-0] Technically elementary (calculus-free), but conceptually challenging problems. Interesting comments and solutions to open-ended problems (e.g., the "Let's Make a Deal" controversy sparked by "Ask Marilyn" vos Savant). RSK

Statistics, P. *Clinical Trials and Statistics.* National Research Council. National Academy Pr, 1993, vii + 45 pp, (P). Proceedings of a 1992 symposium at the National Academy of Sciences, exploring "how statistical research can contribute to gathering increased information from clinical trials and observational studies."

Statistics, T(16-17). *Fundamental Concepts in the Design of Experiments, Fourth Edition.* Charles R. Hicks. Saunders College, 1993, xii + 509 pp, \$49.25. [ISBN 0-03-097710-X] Single-factor and several-factor (factorial, nested hierarchical, nested factorial) experiments. New chapter on Taguchi design proce-

dures in this edition. SAS routines and output accompany many examples. RWJ

Programming, P. *Migrating to Fortran 90*. Jim Kerrigan. O'Reilly & Assoc, 1993, xxvi + 361 pp, \$27.95 (P). [ISBN 1-56592-049-X]

Computer Systems, P. *sendmail*. Bryan Costales, Eric Allman, Neil Rickert. O'Reilly & Assoc, 1993, xxxvi + 792 pp, \$32.95 (P). [ISBN 1-56592-056-2]

Computer Systems, P. *Distributing Applications Across DCE and Windows NT*. Ward Rosenberry, Jim Teague. O'Reilly & Assoc, 1993, xxvi + 274 pp, \$24.95 (P). [ISBN 1-56592-047-3]

Computer Science, P. *System Software and Tools for High Performance Computing Environments*. Eds: Paul Messina, Thomas Sterling. SIAM, 1993, xix + 160 pp, \$5 (P). [ISBN 0-89871-326-9] Report of a 1992 workshop on key issues in developing high-performance computing systems.

Computer Science, T(16-17: 1), L. *The Formal Semantics of Programming Languages: An Introduction*. Glynn Winskel. Found. of Comp. MIT Pr, 1993, xviii + 361 pp, \$40. [ISBN 0-262-23169-7] Up-to-date introduction to programming language semantics. Shows how operational and denotational approaches mesh, and how Gödel's theorem implies impossibility of a fully complete axiomatic semantics. Nicely develops semantics of higher and recursive types, nondeterminism, parallelism. RM

Computer Science, P. *Algorithms and Architectures: Proceedings of the Second NEC Research Symposium*. Ed: T. Ishiguro. SIAM, 1993, x + 282 pp, \$37.50. [ISBN 0-89871-312-9] Highly parallel computing from various points of view. 10 technical papers and synopses of panel discussions from a 1991 symposium in Tsukuba City, Japan.

Applications (Biological Science), P. *Predicting Spatial Effects in Ecological Systems*. Ed: Robert H. Gardner. Lect. on Math. in the Life Sci., V. 23. AMS, 1993, v + 168 pp, \$33 (P). [ISBN 0-8218-1174-6] Proceedings of a symposium at the 1991 Annual Meeting of the AAAS.

Applications (Engineering), P. *Simulation in the Design of Digital Electronic Systems*. John B. Gosling. Electronics Texts for Eng. & Sci. Cambridge Univ Pr, 1993, xv + 273 pp, \$70; \$34.95 (P). [ISBN 0-521-41656-6; 0-521-42672-3]

Applications (Engineering), P. *Lecture Notes in Control and Information Sciences-190: Experimental Robotics II*. Eds: Raja Chatila, Gerd Hirzinger. Springer-Verlag, 1993, xv + 560 pp,

\$95 (P). [ISBN 0-387-19851-2] Proceedings of a 1991 symposium in Toulouse, France.

Applications (Fluid Dynamics), T(17-18: 1). *Mathematical Theory of Incompressible Nonviscous Fluids*. Carlo Marchioro, Mario Pulvirenti. Appl. Math. Sci., V. 96. Springer-Verlag, 1994, xi + 283 pp, \$49. [ISBN 0-387-94044-8] Develops Euler's equation for incompressible nonviscous fluids from viewpoint of mathematical physics. Underlying physical ideas are developed along with rigorous mathematical concepts. Assumes basics of ODE and PDE, measure theory, analytic functions. DS

Applications (Physics), T(16-17: 1, 2), S, P, L. *Inverse Problems in Scattering: An Introduction*. G.M.L. Gladwell. Solid Mechanics & Its Applic., V. 23. Kluwer Academic, 1993, x + 361 pp, \$129. [ISBN 0-7923-2478-1] Aims to display some of the mathematics of one-dimensional inverse scattering problems. First 6 chapters treat layered media; last 4 cover quantum mechanical models. Densely written, numerous exercises. MU

Applications, P. *Mathematical Research in Materials Science: Opportunities and Perspectives*. National Research Council. National Academy Pr, 1993, xi + 129 pp, (P). [ISBN 0-309-04930-X] Surveys past collaborations between mathematics and materials science; then "indicates which particular areas of mathematical sciences research hold the most promise for advancing materials science."

Applications, P. *Transportation and the Mathematical Sciences: The Changing Interaction*. National Research Council. National Academy Pr, 1993, vii + 32 pp, (P). Proceedings of a 1993 symposium at the National Academy of Sciences on planning and operation of surface and air transportation systems.

Applications, P. *Future Directions of Nonlinear Dynamics in Physical and Biological Systems*. Eds: P.L. Christiansen, J.C. Eilbeck, R.D. Parmentier. NATO ASI Ser. B, V. 312. Plenum Pr, 1993, xv + 557 pp, \$139.50. [ISBN 0-306-44562-X] Proceedings of a 1992 NATO Advanced Study Institute in Lyngby, Denmark.

Reviewers

BC: Barry Cipra, St. Olaf; HD: Hung Dinh, Macalester; SG: Steven Galovich, Carleton; TH: Tom Halverson, Macalester; BH: Bruce Hanson, St. Olaf; DH: Deanna Haunsperger, St. Olaf; RWJ: Roger W. Johnson, Carleton; SK: Steve Kennedy, St. Olaf; RSK: Richard S. Kleber, St. Olaf; SM: Steve McKelvey, St. Olaf; RM: Richard Molnar, Macalester; AO: Arnold Ostebee, St. Olaf; DP: David Peifer, St. Olaf; AWR: A. Wayne Roberts, Macalester; KS: Karen Saxe, Macalester; JS: John Schue, Macalester; DS: Dan Schwalbe, Macalester; MU: Milton Ulmer, Carleton; MW: Martha Wallace, St. Olaf; PZ: Paul Zorn, St. Olaf.

EXCURSIONS IN CALCULUS: an Interplay of the Continuous and the Discrete

Robert M. Young

An excellent source of projects for well motivated students. This list of 463 references is a valuable aid for those who wish to dig deeper. —CHOICE

*This excellent book belongs in every school library
and on every mathematics teacher's bookshelf.*
—The Mathematics Teacher

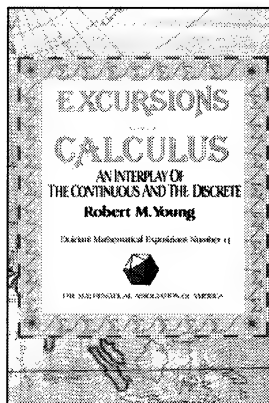
Any book published in the Dolciani Mathematical Expositions series will naturally be measured against the remarkably high standards already well established by the previously published titles. In the present case, however, it is safe to say that this latest addition (the thirteenth in the series) not only easily meets the previously set standards but sets an entirely new standard for future volumes.

—Mathematical Reviews

Printed with eight full-color plates.

The purpose of this book is to explore, within the context of elementary calculus, the rich and elegant interplay that exists between the two main currents of mathematics, the continuous and the discrete. Such fundamental notions in discrete mathematics as induction, recursion, combinatorics, number theory, discrete probability, and the algorithmic point of view as a unifying principle are continually explored as they interact with traditional calculus. The interaction enriches both.

The book is addressed primarily to well-trained calculus students and their teachers, but it can serve as a supplement in a traditional calculus course for anyone who wants to see more.



CONTENTS:

- Infinite Ascent, Infinite Descent: The Principle of Mathematical Induction
- Patterns, Polynomials, and Primes: Three Applications of the Binomial Theorem
- Fibonacci Numbers: Function and Form
- On the Average
- Approximation: from Pi to the Prime Number Theorem
- Infinite Sums: A Potpourri

The problems, taken for the most part from probability, analysis and number theory, are an integral part of the text. Many point the reader toward further excursions. There are over 400 problems presented in this book.

408 pp., 1992, Paperbound

ISBN 0-88385-317-5

List: \$42.00 MAA Member: \$34.00

Catalog Number DOL-13

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

Membership Code

Qty.

Catalog Number

Price

Name _____

Address

City

State Zip Code

Qty.

Catalog Number

Price

Total \$

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No.

Signature _____ Exp. Date _____

Book Scientific Co.

MAIL ORDER & SAVE !! TEXTBOOKS & REFERENCE

Expert Service-Immediate Shipping Worldwide

10% Standard Academic Discount*

SPECIAL OFFER

NEW CUSTOMERS WILL RECEIVE
20% DISCOUNT ON FIRST ORDER**

To order, call **TOLL FREE: 800-621-1220**

Monday to Saturday: 10am-6pm EST

Visa/Mastercard accepted. Catalog available

FAX orders: **(212) 675-4230**

To write: 18 East 16th Street
New York, NY 10003

New York residents only add sales tax

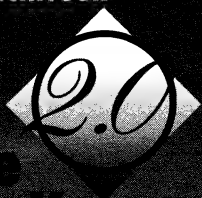
*Minimum two books

** Order limit: Three books. Offer expires Nov. 30th, 1994

Just Released!

THEORIST

For Windows, Macintosh
and Power PC



**Innovative
Solution to Your
Math Problems!**

THEORIST 2.0 is simply the fastest, easiest, most
intuitive package for algebra and graphing ever created.
Experience THEORIST 2.0 and you'll wonder
how you've ever done without it.

CALL 1-800-267-6583

MAPLE

Waterloo Maple Software
(519) 747-2373
info@maplesoft.on.ca

European Headquarters
Waterloo Maple Software GmbH
+49 6221 487 180
100275.163@compuserve.com

Symbolic Mathematics and Graphing 2.0

THEORIST

CONTAINS ESSENTIAL KNOWLEDGE FOR EVERY SCHOOL TEACHER



A FIRST COURSE OF COLLEGIATE MATHEMATICS

by **Joseph B. Dence & Thomas P. Dence**

Orig Ed. 1994 376 pp. \$49.95 ISBN 0-89464-592-7

This text would be useful in courses such as General College Mathematics, Introduction to Mathematical Thinking, and Pre-Calculus.

"It is unjustifiable to make students sit through one more class of the same kind of drudgery that they have seen since their freshman year in high school and thereby deny them exposure to intrinsically more interesting material just because they are not mathematics, science, or engineering majors." - from the Preface

— Call or write for a **FREE** Brochure and Sample Pages —



KRIEGER PUBLISHING COMPANY



P. O. Box 9542, Melbourne, FL 32902-9542
(407) 724-9542 • Direct Order Line (407) 727-7270 • FAX (407) 951-3671

Domestic orders add \$5.00 for first book, \$1.50 each additional for shipping. Foreign shipping costs available upon request.

Visualization in Teaching and Learning Mathematics

**Walter Zimmermann and
Steve Cunningham, Editors**

Buy this book. If you can't buy it, have the library order it. If the library won't order it, ask to borrow a copy from a friend. But do read this book.

—*The Mathematics Teacher*

High school, community college, and university teachers who use or are interested in using graphics to teach calculus, deductive reasoning, functions, geometry, or statistics will find valuable ideas for teaching... A must for every college or university library with a mathematics department.—CHOICE

The twenty papers in this book give an overview of research, analysis, practical experience, and informed opinion about the role of visualization in teaching and learning mathematics, especially at the undergraduate level. Visualization in its broadest sense is as old as mathematics, but

progress in computer graphics has generated a renaissance of interest in visual representations and visual thinking in mathematics.

230 pp., Paperbound, 1991

ISBN 0-88385-071-0

List: \$24.00

Catalog Number NTE-19

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

**No matter how you
express it, it still means
DERIVE[®] is half price.**

DERIVE 

The *DERIVE A Mathematical Assistant* program lets you express yourself symbolically, numerically and graphically, from algebra through calculus, with vectors and matrices too—all displayed with accepted math notation, or 2D and 3D plotting. *DERIVE* is also easy to use and easy to read, thanks to a friendly, menu-driven interface and split or

overlay windows that can display both algebra and plotting simultaneously. Better still, *DERIVE* has been praised for the accuracy and exactness of its solutions. But, best of all the suggested retail price is now only \$125. Which means *DERIVE* is now half price, no matter how you express it.

System requirements

DERIVE: MS-DOS 2.1 or later, 512K RAM, and one 3 $\frac{1}{2}$ " disk drive. Suggested retail price now **\$125 (Half off!)**.

DERIVE ROM card: Hewlett Packard 95LX & 100LX Palmtop, or other PC compatible ROM card computer. Suggested retail price now **\$125!**

DERIVE XM (eXtended Memory): 386 or 486 PC compatible with at least 2MB of *extended* memory. Suggested list price now \$250!

DERIVE is a registered trademark of Soft Warehouse, Inc.

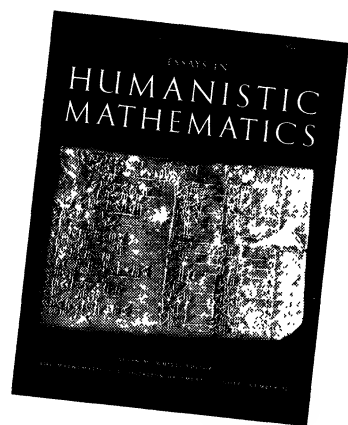


Soft Warehouse
HONOLULU • HAWAII

Soft Warehouse, Inc. • 3660 Waialae Ave.
Ste. 304 • Honolulu, HI, USA 96816-3236
Ph: (808) 734-5801 • Fax: (808) 735-1105

Essays in Humanistic Mathematics

Alvin White, Editor



A dazzling array of essayists reveal humanistic mathematics in this volume, and in so doing go beyond the facts, formulas, and algorithms that most students associate with mathematics to a presentation of mathematics as an intellectual discipline with a human perspective and a significant history. Humanistic mathematics challenges dogmatic teaching styles that expect students to parrot the lecturer. It demands creativity from both the teacher and student.

Teaching mathematics humanistically seeks to place the student more centrally in the position of inquirer than is generally the case, while at the same time acknowledging the emotional climate of the activity of learning mathematics. This type of teaching encourages students to learn from each other and to better understand mathematics as socially constructed knowledge, rather than as an arbitrary discipline.

Teaching humanistic mathematics brings the focus less upon the nature of the teaching and learning environment and more upon the need to reconstruct the curriculum and the discipline of mathematics itself. This reconstruction relates mathematical discoveries to personal courage, discovery to verification, mathematics to science, truth to utility, and

mathematics to the culture in which it is embedded.

The humanistic mathematics movement, which began as the personal vision of a few, has now become a major part of mathematical culture. What was viewed with skepticism is now accepted and expected. Humanistic mathematics is not a new discovery. It is a recent rediscovery of ideas that go back to Plato. It has provided a vocabulary for previously unarticulated concepts and approaches.

The essays in this volume illustrate and help to define humanistic mathematics. The variety and scope indicate the richness and fruitfulness of the concept. Although each essay is independent, a sense of unity emerges. A glimpse at the table of contents will give you an idea of the excitement and range of the ideas presented.

212 pp., Paperbound, 1993

ISBN 0-88385-089-3

List: \$24.00

Catalog Number NTE-32

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Name _____

Address _____

City _____

State ____ Zip Code _____

Qty.	Catalog Number	Price
------	----------------	-------

_____	_____	_____
-------	-------	-------

		Total \$ _____
--	--	----------------

Payment ☐ Check ☐ VISA ☐ MASTERCARD

Credit Card No. _____

Signature _____

Exp. Date _____

From Zero to Infinity

Constance Reid
Fourth Edition

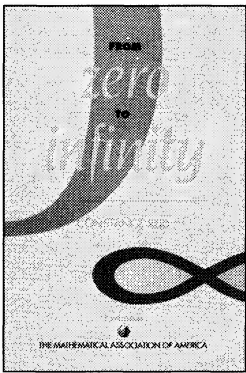
In a delightful way, it well serves the purpose of the author—to get others interested in the fascinating theory of numbers. —The Mathematics Teacher

This is a book one can read and enjoy from cover to cover in one sitting... It is highly recommended for all. —Crux Mathematicorum

The 1992 version of a graceful book, justly praised for almost 40 years... Any closely attentive reader who knows schoolroom arithmetic and the use of exponents is prepared for these pages, and few outside the profession will find the material too familiar. —Scientific American

From Zero to Infinity has dazzled readers with its freshness and clarity since being published in 1955. This book shows how interesting the everyday natural numbers 0, 1, 2, 3, ... have been for over two thousand years, and still are today. It combines the mathematics and the history of number theory with descriptions of the mystique that has on occasion surrounded numbers even among great mathematicians.

Each chapter takes one of the ten digits as a starting point. In some cases, as with 0 and 1, the numbers are in themselves special and unique. In other cases, as with 4 (the first square after the trivial 1²), or 6 (the first perfect number), each digit serves to introduce an infinite series



of very interesting numbers and the very interesting mathematical questions that arise in connection with them.

200 pp., Paperbound, 1992

ISBN-0-88385-505-4

List: \$22.00 MAA Member: \$16.50

Catalog Number ZTI

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Foreign Orders Please add \$3.00 per item ordered to cover postage and handling fees. The order will be sent via surface mail. If you want your order sent by air, we will be happy to send you a proforma invoice for your order.

	Qty.	Catalog Number	Price
Membership Code _____			
Name _____			Total \$ _____
Address _____			Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
City _____			Credit Card No. _____
State _____ Zip Code _____			Signature _____
			Exp. Date _____

PRINCIPLES *of* SOUND RETIREMENT INVESTING

“CREF Stock Account ★★★★★★”
“CREF Bond Market Account . . . ★★★★★★”
“CREF Social Choice Account . . . ★★★★★★”

—Morningstar

ISN'T IT NICE WHEN THE EXPERTS DISCOVER SOMETHING YOU'VE KNOWN ALL ALONG.

Over 1.6 million people in education and research know that choosing TIAA-CREF was a smart move. And now everyone else does too. Because Morningstar—one of the nation's leading sources of variable annuity and mutual fund information—has some stellar things to say about our retirement investment accounts.

“This comfortable combination of risk and return has earned the CREF Stock Account a five-star rating.”*

After studying CREF's performance history, Morningstar gave five-stars—its highest rating—to both the CREF Stock and Bond Market Accounts, and an impressive four-stars to the CREF Social Choice Account.** In fact, the CREF Stock Account was singled out as having “...one of the best 10-year records among variable

annuities.”*** Of course, past performance is no guarantee of future results.

“...CREF is far and away the cheapest variable annuity out there.”

Morningstar also called attention to CREF's “...rock-bottom” fees—something that can really add to the size of your nest egg down the road.

What's more, TIAA's traditional annuity—which offers guaranteed principal and interest plus the opportunity for dividends—was cited as having the highest fixed account interest rate among all annuities in its class.

We're happy to accept Morningstar's glowing ratings. But nice as it is to focus on stars, we'll keep focusing on something more down-to-earth: building the financial future you want and deserve.

For more information about our Morningstar ratings or TIAA-CREF just call 1 800 842-2776.



**Ensuring the future
for those who shape it.™**

*Source: Morningstar's Comprehensive Variable Annuity/Life Performance Report January, 1994

**Source: Morningstar Inc. for periods ending March 31, 1994. Morningstar is an independent service that rates mutual funds and variable annuities on the basis of risk-adjusted performance. These ratings are subject to change every month. The top 10% of funds in each class receive five stars, the following 22.5% receive four stars.

***Among the variable annuity accounts ranked by Morningstar: the CREF Stock Account was 1 of 12 growth-and-income accounts with 10 years of performance. Morningstar ranks the performance of a variable annuity account relative to its investment class based on total returns. CREF certificates are distributed by TIAA-CREF Individual and Institutional Services. For more complete information, including charges and expenses, call 1 800 842-2776 for a prospectus. Read the prospectus carefully before you invest or send money.

How Many Candles Were On Your Cake The Last Time You Thought About Buying Insurance?

Face it—it's been a long time. Styles have changed. So has your family, maybe even your job. And most likely, the insurance you bought then isn't enough to cover your family today. That's why you need coverage that you can easily update as your life changes—MAA Group Insurance Program.

We Understand You.

Finding an insurance program that's right for you isn't easy. But as a member of MAA, you don't have to go through the difficult task of looking for the right plans—we've done that work for you. What's more, the program is constantly being evaluated to better meet the needs of our members.

We're Flexible.

Updating your insurance doesn't have



to be a hassle. With our plans, as your needs change, so can your coverage. Insurance through your association is designed to grow with you—it even moves with you when you change jobs.

We're Affordable.

We offer members the additional benefit of reasonable rates, negotiated using our group purchasing power. Call 1 800 424-9883 (in Washington, DC, (202) 457-6820) between 8:30 a.m. and 5:30 p.m. Eastern Time for more information about these insurance plans offered through MAA:

Term Life • Disability Income Protection • Comprehensive HealthCare • Excess Major Medical • In-Hospital • High-Limit Accident

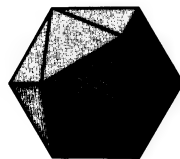
MAA Insurance

Designed for the way you live today.
And tomorrow.

This Plan is Administered by Seabury & Smith.

The American Mathematical Monthly

Volume 101 Number 8 / OCTOBER 1994
(ISSN 0002-9890)



Contents

ARTICLES

Behind the Scenes of a Random Dot Stereogram / MARIA S. TERRELL
and ROBERT E. TERRELL 715

The Fifty-Fourth William Lowell Putnam Mathematical Competition /
LEONARD F. KLOSINSKI, GERALD L. ALEXANDERSON,
and LOREN C. LARSON 725

Literacy in the Language of Mathematics / JAMES O. BULLOCK 735

Fractional and Trigonometric Expressions for Matrices /
GORO SHIMURA 744

Noether Lasker Primary Decomposition Revisited /
BARBARA L. OSOFSKY 759

Elementary Infinite Sources of Non-Unique Factorization Rings / S. STEIN
and S. SZABÓ 769

Apropos Two Notes on Notation / ANTAL E. FEKETE 771

FEATURES

COMMENTS 714

NOTES

The Coin Exchange Problem for Arithmetic Progressions /
AMITABHA TRIPATHI 779

Congruence of Triangles / LEONARD GILLMAN 782

A Short Elementary Proof of the Mohr-Mascheroni Theorem /
NORBERT HUNGERBÜHLER 784

UNSOLVED PROBLEMS

Which Triangles Are Plane Sections of Regular Tetrahedra? /
FOLKE ERIKSSON 788

THE AUTHORS 790

PROBLEMS AND SOLUTIONS 792

REVIEWS

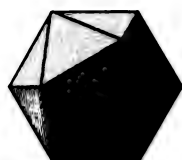
Brouwer's Intuitionism. By Walter P. van Stigt / C. SMORYŃSKI 799

TELEGRAPHIC REVIEWS 803

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street N.W.



The American Mathematical Monthly



Volume 101, Number 9 / NOVEMBER 1994



AN OFFICIAL PUBLICATION OF THE MATHEMATICAL ASSOCIATION OF AMERICA

NOTICE TO AUTHORS

The *Monthly* publishes articles, notes, and other features about mathematics and the profession. The readership of the *Monthly* is intended to include everybody who is mathematically inclined, including of course professional mathematicians and students of mathematics at all collegiate levels. While no single article or feature is likely to appeal to everyone, material should interest and be accessible to a large number of readers. This is the most important criterion for acceptance.

Articles may be expositions of old results or presentations of new ones. They may concern all of mathematics or one small area, a broad development or a single application, historical reminiscences or one important event. While some articles may contain the author's new research, the novelty of material and generality of the results is far less important than the clarity of exposition and general interest. Discussing one illuminating case of a well known result is far better than providing all the details of an obscure but new proposition. Articles in the *Monthly* are supposed to inform and to entertain; they are meant to be read rather than archived.

Notes are short and possibly informal articles. A note may concern a clever new proof of an old theorem, a novel way to present tired material, or a lively discussion of a philosophical (but still mathematical) issue. Also, any topic is suitable, so long as it is related to mathematics. Because a note is short, the first few sentences are the most important part: They should explain the purpose and invite the reader in. Photographs or diagrams often will attract the reader's attention.

All articles and notes should be sent to the editor:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Please send 3 copies, typewritten on only one side of the paper. Illustrations should be carefully drawn on separate sheets of paper in black ink; the original should be without lettering and two copies should have appropriate captions and lettering indicated.

Proposed problems or solutions should be sent to:

RICHARD BUMBAY,
P.O. Box 10971
New Brunswick, NJ 08906-0971.

Please send 2 copies of all material, typewritten if possible.

Letters to the Editor, both for publication and for private reading, should be sent to the Editor at the address given above. Comments, including criticisms, are welcome, as are all suggestions for making the *Monthly* a lively, entertaining, and informative journal.

Cover Photo: Georg Cantor in his middle years, likely taken between 1880 and 1894. The photograph was obtained through the courtesy of the late Wilhelm Stahl and Ivor Grattan-Guinness.

EDITOR:

JOHN H. EWING

ASSOCIATE EDITORS:

PETER BORWEIN	FRED KOCHMAN
RICHARD BUMBAY	CATHERINE MCGEOCH
DENNIS DETURCK	RICHARD NOWAKOWSKI
UNDERWOOD DUDLEY	ARNOLD OSTEBEE
JOHN DUNCAN	LEE RUBEL
JOAN FERRINI-MUNDY	ABE SHENITZER
JOSEPH GALLIAN	LYNN STEEN
STEVEN GALOVICH	STAN WAGON
RICHARD GUY	DOUGLAS WEST
DARRELL HAILE	HERBERT WILF
PAUL HALMOS	SANDY ZABELL
JOAN HUTCHINSON	PAUL ZORN

EDITORIAL ASSISTANT:

MISTY CUMMINGS

STAFF ARTIST:

MIKE CAGLE

QUOTE MASTER:

MARK WOODARD

Reprint permission:
MARCIA P. SWARD, Executive Director

Advertising Correspondence:
Ms. ELAINE PEDREIRA, Advertising Manager

Subscription correspondence, change of address,
and other inquiries:
Membership / Subscriptions Department

All at the address:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036.

Microfilm Editions: University Microfilms International,
Serial Bid coordinator, 300 North Zeeb Road, Ann
Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT. Copyrighted by the Mathematical Association of America (Incorporated), 1995, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source. Second class postage paid at Washington, DC, and additional mailing offices. **Postmaster:** Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.



Contents

ARTICLES

- Georg Cantor and Transcendental Numbers / ROBERT GRAY 819
The 500th Anniversary of the Sharing Problem (The Oldest Problem
in the Theory of Probability) / MILTON SOBEL
and KRZYSZTOF FRANKOWSKI 833
What Is Teaching? / PAUL R. HALMOS 848
What Is Wrong with the Definition of dy/dx ? / HUGH THURSTON 855
A Stochastic Approach to the Gamma Function / LOUIS GORDON 858
Arrangements and Topological Planes / JACOB E. GOODMAN,
RICHARD POLLACK, RAPHAEL WENGER,
and TUDOR ZAMFIRESCU 866
An Application of Fourier Series to the Most Significant Digit Problem /
JEFF BOYLE 879
-

FEATURES

COMMENTS. 818

NOTES

- Cross Product Identities in Arbitrary Dimension /
ANDREW DITTMER 887
A Non-Constant, Continuous Function on the Plane Whose Integral
on Every Line Is Zero / D. H. ARMITAGE 892
Chu's 1303 Identity Implies Bombieri's 1990 Norm-Inequality
(Via an Identity of Beauzamy and Dégot) /
DORON ZEILBERGER 894

THE COMPUTER SCIENCE SAMPLER

- How to Stay Competitive / CATHERINE C. McGEOCH 897

THE EVOLUTION OF...

- On the Calculus of Variations and Its Major Influences
on the Mathematics of the First Half of Our Century. Part II /
ERWIN KREYSZIG 902

THE AUTHORS 909

PROBLEMS AND SOLUTIONS 911

REVIEWS

- Fleeting Footsteps: Tracing the Conception of Arithmetic and Algebra
in Ancient China.* By Lam Lay Yong and Ang Tian Se /
FRANK SWETZ 921

TELEGRAPHIC REVIEWS 924

COMMENTS

In the death of Professor Eliakim Hastings Moore on December 30, 1932, the University of Chicago lost one of its most inspiring leaders and the world ...of mathematicians lost one of its most gifted members. ... Professor Moore was one of the youngest of this group of men [at Chicago] but his dauntless spirit, his boundless enthusiasm, his scientific integrity, and his unsurpassed logical insight soon established an influence which has extended far and wide and will continue to be felt wherever students of mathematics are seeking to discover the mainspring of their inspiration.

Looking back over a period of forty-one years ... this writer ... now wishes to enumerate certain outstanding characteristics of this truly great man and, if possible, to set them forth with sufficient emphasis ... to attract the emulation of the on-coming generation of students of mathematics.

First, and perhaps most important, was Professor Moore's exemplification of *intellectual honesty*. It was impossible for him to accept as established any proposition whose proof was in doubt... There was no place for camouflage of any sort in his mental make-up. A striking and critical test of this characteristic occurred at the opening meeting of the Mathematical Club in one of the early summer quarters. Moore was presenting a paper on a highly technical topic to a large gathering of faculty and graduate students from all parts of the country. When half way through he discovered what seemed to be an error (though probably no one else in the room observed it). He stopped and re-examined the doubtful step for several minutes and then, convinced of the error, he abruptly dismissed the meeting — to the astonishment of most of the audience. It was an evidence of intellectual *courage* as well as *honesty* and doubtless won for him the supreme admiration of every person in the group — an admiration which was in no wise diminished, but rather increased, when at a later meeting he announced that after all he had been able to prove the step to be correct.

A second outstanding characteristic was the *intellectual freedom* which he claimed for himself and which he cheerfully granted to all others. He expected every man on the staff to do his full duty but he did not follow him up or visit his classroom. He trusted, rather, to the freedom based upon individual responsibility and, on the whole, the results were far more satisfactory than any system of espionage could possibly have been. There were many departmental open discussions where every instructor's views could be fully weighed and evaluated in the light of all other considerations; but in the end individual freedom was accorded to each one for working out the general plan. If in any case the results were unsatisfactory, it was up to the instructor to discover the trouble and to find the remedy. To this acid test of *freedom*, Professor Moore subjected himself on equal terms with all his instructors, even to the extreme test, on occasion, of admitting his own failure and seeking the remedy from the counsel of others. ...

H. E. Slaught
Monthly 40(1933), 191-195.

Georg Cantor and Transcendental Numbers

Robert Gray

1. INTRODUCTION. Conflicting statements have been made about Cantor's proof of the existence of transcendental numbers. For example, consider the following statements:

The contrast between the methods of Liouville and Cantor is striking, and these methods provide excellent illustrations of two vastly different approaches toward proving the *existence* of mathematical objects. Liouville's is purely *constructive*; Cantor's is purely *existential*.

—Mark Kac and Stanislaw M. Ulam [20, p. 13]

Sometimes people have gone on to say that Cantor's method is not constructive, and cannot yield an explicit transcendental number. The words of E. T. Bell on page 569 of his *Men of Mathematics* are typical:

The most remarkable thing about Cantor's proof is that it provides no means whereby a single one of the transcendentals can be constructed.

This is not true. Cantor's idea can be used so as to yield an utterly explicit transcendental number.

—I. N. Herstein and I. Kaplansky [19, p. 238]

Kac and Ulam are comparing Cantor's method with Liouville's earlier construction of transcendentals [24, 25]—they find Cantor's method to be non-constructive. But Herstein and Kaplansky insist that Cantor's method is constructive, that it can produce a transcendental. A few lines later, they assert that this transcendental “is as well determined a number as e or π .”

After reading the statements by Kac and Ulam, and Herstein and Kaplansky, we decided to study Cantor's work and how it has been presented. This article contains the results of our study. We begin by analyzing Cantor's original articles, his 1874 article that contains his first proof and his 1891 article that contains his diagonal proof. Our analysis will show that Cantor's methods lead to computer programs that generate transcendentals, and it will also determine which transcendentals are generated by the diagonal method. Next we will examine the history behind Cantor's first proof. Finally, we will consider how some commonly-held views about mathematics and its history have affected the interpretation of Cantor's work.

2. CANTOR'S FIRST PROOF. In 1874, Cantor published his first proof of the existence of transcendentals in an article titled “On a Property of the Collection of All Real Algebraic Numbers” [3, 5]. Cantor begins his article by defining the

algebraic reals and introducing the notation: (ω) for the collection of all algebraic reals, and (ν) for the collection of all natural numbers. Next he states the property mentioned in the article's title; namely, that the collection (ω) can be placed into a one-to-one correspondence with the collection (ν) , or equivalently:

... the collection (ω) can be thought of in the form of an infinite sequence:

$$(2.) \qquad \qquad \qquad \omega_1, \omega_2, \dots, \omega_\nu, \dots$$

which is ordered by a law and in which all individuals of (ω) appear, each of them being located at a fixed place in (2.) that is given by the accompanying index.

Cantor states that this property of the algebraic reals will be proved in Section 1 of his article, and then he outlines the rest of the article:

To give an application of this property of the collection of all real algebraic numbers, I supplement Section 1 with Section 2, in which I show that when given an arbitrary sequence of real numbers of the form (2.), one can determine, in any given interval $(\alpha \cdots \beta)$, numbers η that are not contained in (2.). Combining the contents of both sections thus gives a new proof of the theorem first demonstrated by Liouville: In every given interval $(\alpha \cdots \beta)$, there are infinitely many transcendentals, that is, numbers that are not algebraic reals. Furthermore, the theorem in Section 2 presents itself as the reason why collections of real numbers forming a so-called continuum (such as, all the real numbers which are ≥ 0 and ≤ 1), cannot correspond one-to-one with the collection (ν) ; thus I have found the clear difference between a so-called continuum and a collection like the totality of all real algebraic numbers.

To appreciate the structure of Cantor's article, we number his theorems and corollaries:

Theorem 1. *The collection of all algebraic reals can be written as an infinite sequence.*

Theorem 2. *Given any sequence of real numbers and any interval $[\alpha, \beta]$, one can determine a number η in $[\alpha, \beta]$ that does not belong to the sequence. Hence, one can determine infinitely many such numbers η in $[\alpha, \beta]$. (We have used the modern notation $[\alpha, \beta]$ rather than Cantor's notation $(\alpha \cdots \beta)$.)*

Corollary 1. *In any given interval $[\alpha, \beta]$, there are infinitely many transcendental reals.*

Corollary 2. *The real numbers cannot be written as an infinite sequence. That is, they cannot be put into a one-to-one correspondence with the natural numbers.*

Observe the flow of reasoning: Cantor's second theorem holds for *any* sequence of reals. By applying his theorem to the sequence of algebraic reals, Cantor obtains transcendentals. By applying it to any sequence that allegedly enumerates the reals, he obtains a contradiction—so no such enumerating sequence can exist. Kac and Ulam reason differently [20, p. 12–13]. They prove Theorem 1 and then Corollary 2. By combining these results, they obtain a non-constructive proof of the existence of transcendentals.

Cantor's theorems are worded constructively, but are they proved constructively? Since Cantor's original proof of Theorem 2 is not commonly known, we present it before answering this question.

Proof of Theorem 2: Recall that we have been given an interval $[\alpha, \beta]$ and a sequence of real numbers ω_n . We must find an η in $[\alpha, \beta]$ that does not belong to this sequence. Cantor assumes that the members of the sequence are distinct; to handle an arbitrary sequence, we can either eliminate duplicates from the sequence or modify his proof to handle arbitrary sequences.

Cantor begins his proof by finding the first two numbers in the given sequence that belong to $[\alpha, \beta]$. Denote the smaller of these numbers by α_1 and the larger by β_1 . Now form the interval $[\alpha_1, \beta_1]$, and locate the first two numbers in the sequence that belong to $[\alpha_1, \beta_1]$. Denote the smaller of these numbers by α_2 and the larger by β_2 . Then form the interval $[\alpha_2, \beta_2]$, and continue this procedure of generating intervals.

We have two cases: Cantor's procedure yields finitely many intervals $[\alpha_n, \beta_n]$ or infinitely many such intervals. In the first case, let $[\alpha_N, \beta_N]$ be the last interval generated. Since there can be at most one ω_k in $[\alpha_N, \beta_N]$, any η in the interval besides this ω_k and the endpoints of the interval will satisfy the conclusion of the theorem. In the second case, let $\alpha_\infty = \lim_{n \rightarrow \infty} \alpha_n$ and $\beta_\infty = \lim_{n \rightarrow \infty} \beta_n$. These limits exist since the α_n 's form an increasing sequence that is bounded from above, and the β_n 's form a decreasing sequence that is bounded from below.

The second case breaks into two cases: Either $\alpha_\infty = \beta_\infty$ or $\alpha_\infty < \beta_\infty$. At this point in his proof, Cantor notes that $\alpha_\infty = \beta_\infty$ holds for the sequence of algebraic reals [3, p. 261; 5, p. 308–309]. So Cantor not only applies his theorem to the sequence of algebraic reals, but he also analyzes how his proof handles this particular sequence.

To complete the proof, we must produce a suitable η for the two remaining cases. In the case where $\alpha_\infty = \beta_\infty$, let η be this common limit. Note that η cannot be a member of the given sequence since for every k , ω_k does not belong to $[\alpha_{k+1}, \beta_{k+1}]$. In the case where $\alpha_\infty < \beta_\infty$, let η be any number in the interval $[\alpha_\infty, \beta_\infty]$.

Cantor's proof is constructive—he uses the given sequence ω_n to define the sequences α_n and β_n , breaks his argument into three cases depending on the behavior of these sequences, and constructs a suitable η for each case.* If the sequence ω_n contains all the algebraic reals, then α_n and β_n are converging nested sequences whose common limit η is a transcendental number.

Perhaps the most convincing way to show that Cantor's argument produces a transcendental is by computing one. Using the methods in Cantor's proof, we have written a computer program that generates the digits of a transcendental in the interval $(0, 1)$. Output from our program is given in Figure 1. Our program generates the sequence ω_n by enumerating the polynomials with integer coefficients and approximating their roots. We approximate roots by using Sturm's theorem and Horner's method [31, p. 138–156]. (For a precise description of the ω_n sequence generated by our program, see the appendix below.)

*We call a proof “constructive” if it constructs an object using methods acceptable to most mathematicians. For a proof of Cantor's Theorem 2 that meets the demands of constructive mathematicians, see [2, p. 27].

Approximation	Associated Polynomial
$\alpha_1 = \omega_1 = .50$ $\beta_1 = \omega_2 = .61\dots$	$2x - 1$ $x^2 + x - 1$
$\alpha_2 = \omega_{12} = .561\dots$ $\beta_2 = \omega_{19} = .577\dots$	$x^2 + 3x - 2$ $3x^2 - 1$
$\alpha_3 = \omega_{41} = .569\dots$ $\beta_3 = \omega_{66} = .574\dots$	$x^3 - x^2 + 2x - 1$ $x^3 + 2x^2 + 2x - 2$
$\alpha_4 = \omega_{87} = .57318\dots$ $\beta_4 = \omega_{359} = .57347\dots$	$2x^3 - 2x^2 - 3x + 2$ $2x^3 + x^2 + 4x - 3$
$\alpha_5 = \omega_{5539} = .573402\dots$ $\beta_5 = \omega_{2159} = .573416\dots$	$4x^4 - 3x^3 + 3x^2 + 2x - 2$ $2x^4 - 4x^3 - 4x^2 - 2x + 3$
$\alpha_6 = \omega_{156510} = .5734104\dots$ $\beta_6 = \omega_{144803} = .5734122\dots$	$3x^5 + 5x^4 - x^3 + 4x^2 + 2x - 3$ $3x^5 + 3x^4 - 2x^3 - 3x^2 - 2x + 2$
$\alpha_7 = \omega_{1406370} = .57341146\dots$ $\beta_7 = \omega_{1057887} = .57341183\dots$	$x^6 - x^5 + 2x^4 + 3x^3 - x^2 + x - 1$ $x^6 - 4x^5 - x^4 + 5x^3 + 2x^2 + 3x - 3$

Figure 1. Generating a transcendental using Cantor’s 1874 method.

While generating the ω_n sequence, our program also generates the sequences α_n and β_n , which approximate our transcendental η . Figure 1 shows the first seven members of α_n and β_n . As Cantor points out in his proof, these sequences have a common limit—so our program will produce closer approximations. In fact, we can calculate a bound on the number of algebraic reals that need examining in order to find an α_n and β_n that approximate η to within $1/k$. This calculation requires a look at our enumeration of the algebraic reals.

For ease of programming, we use a different enumeration of the algebraic reals than Cantor’s. Cantor enumerates the polynomials with integer coefficients by their *height*, where the height of the polynomial $a_0x^k + \dots + a_k$ is $k - 1 + |a_0| + \dots + |a_k|$, and then he enumerates the roots of these polynomials. We enumerate polynomials by their *size*, which for the polynomial just mentioned is $\max(k, |a_0|, \dots, |a_k|)$. Polynomials of the same size are ordered by treating them as $(k + 1)$ -digit numbers whose digits range from $-k$ to k . For example, the polynomials $20x - 1$ and $x^{20} - 1$ are both size 20 and the first precedes the second in this ordering. Our enumeration of polynomials generates an enumeration of their roots; when a polynomial has more than one real root, we enumerate its roots in numerical order.

To obtain an α_n and β_n such that $\beta_n - \alpha_n \leq 1/k$, we first enumerate those roots of polynomials of size $2k$ or less that are in the interval $(0, 1)$. Applying the procedure in Cantor’s proof to this enumeration, we obtain the finite sequence $\alpha_1, \beta_1, \alpha_2, \beta_2, \dots, \alpha_n, \beta_n$. Now if $\beta_n - \alpha_n > 1/k$, then we would have $\alpha_n < j/(2k) < (j + 1)/(2k) < \beta_n$ for some j between 1 and $2k - 2$. Since $j/(2k)$ and $(j + 1)/(2k)$ are roots of polynomials of size $2k$, we have found two roots of our enumeration between α_n and β_n . But Cantor’s procedure allows at most one of these roots to be between α_n and β_n . Hence, we must have $\beta_n - \alpha_n \leq 1/k$. Since there are $(4k + 1)^{2k+1} - (4k + 1)$ polynomials of size $2k$ or less, and since each one has at most $2k$ roots, we need to examine at most $2k[(4k + 1)^{2k+1} - (4k + 1)]$ algebraic reals in order to approximate our transcendental η to within $1/k$.

This simple argument produces a poor bound. Figure 1 shows that we do not need to enumerate polynomials of size 200 to obtain approximations differing by less than $1/100$. Nevertheless, our computer program generates digits inefficiently

—asymptotically, it takes at least $O(2^{\sqrt[3]{n}})$ steps to generate n digits of our transcendental number [18]. Any program requiring this many steps is regarded as inefficient by computer scientists [16, p. 6–9].

Cantor's proof leads to a computer program that generates a transcendental, but a program is not necessary for understanding his article. The constructive nature of Cantor's article is clear from the wording and proof of Theorem 2. This theorem separates the constructive part of his article from the proof-by-contradiction needed to establish the uncountability of the set of reals. Since we will be referring to Theorem 2 throughout our article, we shall give it a name—*Cantor's theorem on real sequences*.

We are far from the first to point out that Cantor's article is constructive. In 1930, Fraenkel stated that the method in this article is “a method that incidentally, contrary to a widespread interpretation, is fundamentally constructive and not merely existential” [15, p. 237].

Exercise. The sequence $1/2, 1/3, 2/3, 1/4, 2/4, 3/4, \dots$ contains all the rationals belonging to $(0, 1)$. Apply the algorithm in Cantor's proof to this sequence to generate the digits of an irrational. If you are using pencil and paper, just compute $\alpha_1, \beta_1, \alpha_2$, and β_2 .

3. CANTOR'S DIAGONAL PROOF. We now turn to Cantor's 1891 article [9], which contains his well-known diagonal proof. Cantor begins by discussing his 1874 article. He points out that it contains a proof of the theorem: There are infinite sets that cannot be put into one-to-one correspondence with the set of positive integers. Then he asserts that this theorem has a much simpler proof than the one given in 1874. His new proof uses the set M of elements of the form $E = (x_1, x_2, \dots, x_\nu, \dots)$, where each x_ν is either m or w . Cantor states that M is uncountable, and notes that this result is implied by the following theorem:

If $E_1, E_2, \dots, E_\nu, \dots$ is any simply infinite sequence of elements of the set M , then there is always an element E_0 of M which corresponds to no E_ν .

Cantor proves his theorem by using the diagonal method to construct E_0 . Note that, once again, Cantor states a theorem that separates the constructive content of his work from the proof-by-contradiction needed to establish uncountability.

By introducing sequences of abstract symbols, Cantor shows that the phenomenon of uncountability does not depend on properties of the real numbers, such as the existence of limits for bounded increasing (or decreasing) sequences of reals. Thus, Cantor shows that uncountability is a fundamental phenomenon of set theory. Also, his 1874 theorem on real sequences follows easily from his new theorem. Take any sequence of real numbers and expand its members into their binary representations. (A real of the form $m/2^n$ must be expanded into both of its binary representations.) This gives us a sequence of binary representations. Apply Cantor's new theorem to obtain the binary representation of a real number that does not belong to the original sequence.

Cantor's diagonal method is simpler than his earlier nesting method, and it generates transcendentals much more efficiently. The diagonal method can generate n digits of a transcendental in $O(n^2 \log^2 n \log \log n)$ steps [18]. Algorithms requiring less than $O(n^3)$ steps are considered practical by computer scientists [16, p. 9].

Figure 2 contains output from a computer program that uses Cantor’s diagonal method to generate the digits of a transcendental in $(0, 1)$. Our program generates the same ω_n sequence as our program in Section 2. It uses this sequence to generate the digits of a diagonal number as follows: Let d be the n th digit of ω_n . Our program sets the n th digit of our diagonal number to $d + 1$ unless d is 9; in this case, it sets the n th digit to 0. Cantor’s diagonal argument guarantees that the decimal representation of our diagonal number differs from the representations we use for the algebraic reals. But as Figure 2 shows, our program does not generate both representations of fractions such as $1/2$. Hence, we cannot conclude that our diagonal number is transcendental until we show that it differs from all fractions having two decimal representations. Our diagonal number does differ from these fractions because its decimal expansion contains infinitely many 2’s—the decimal expansions of $1/9, 1/90, 1/900, \dots$ generate 2’s on the diagonal.

Note that our diagonal number can be written as:

$$\sum_{n=1}^{\infty} \frac{\text{rem}(\lfloor 10^n \cdot \omega_n \rfloor + 1, 10)}{10^n}$$

where ω_n is the n th member of our sequence of algebraic reals; $\text{rem}(m, n)$ is the remainder left after dividing m by n ; and $\lfloor x \rfloor$ is the *floor* of x (the largest integer equal to or less than x).

Exercise. Write the rationals in $(0, 1)$ as a sequence: $1/2, 1/3, 2/3, 1/4, 2/4, 3/4, \dots$. By applying the diagonal method to this sequence, generate an irrational number in $(0, 1)$. Compute the first 10 digits of this number, compute the 25th

Algebraic Real	Associated Polynomial	Approximation to Transcendental
$\omega_1 = .5$	$2x - 1$.6
$\omega_2 = .61\dots$	$x^2 + x - 1$.62
$\omega_3 = .732\dots$	$x^2 + 2x - 2$.623
$\omega_4 = .4142\dots$	$x^2 + 2x - 1$.6233
$\omega_5 = .70710\dots$	$2x^2 - 1$.62331
$\omega_6 = .780776\dots$	$2x^2 + x - 2$.623317
$\omega_7 = .3660254\dots$	$2x^2 + 2x - 1$.6233175
$\omega_8 = .66666666\dots$	$3x - 2$.62331757
$\omega_9 = .33333333\dots$	$3x - 1$.623317574
$\omega_{10} = .3819660112\dots$	$x^2 - 3x + 1$.6233175743
$\omega_{11} = .79128784747\dots$	$x^2 + 3x - 3$.62331757438
$\omega_{12} = .561552812808\dots$	$x^2 + 3x - 2$.623317574389
$\omega_{13} = .3027756377319\dots$	$x^2 + 3x - 1$.6233175743890
$\omega_{14} = .8228756553229\dots$	$2x^2 + 2x - 3$.62331757438900
$\omega_{15} = .686140661634507\dots$	$2x^2 + 3x - 3$.623317574389008

Figure 2. Generating a transcendental using Cantor’s diagonal method.

digit. Verify that this number can be written as:

$$\sum_{n=2}^{\infty} \sum_{m=1}^{n-1} \frac{\text{rem}\left(\left\lfloor 10^{f(m,n)} \cdot \frac{m}{n} \right\rfloor + 1, 10\right)}{10^{f(m,n)}}$$

where $f(m, n) = (n - 2)(n - 1)/2 + m$.

4. ALL TRANSCENDENTALS LIVE ON DIAGONALS. As we have seen, Cantor's diagonal method does construct transcendentals—but which ones? To answer this question, we first observe that the digits generated by the diagonal method depend on the enumeration of algebraic reals we use. Different enumerations usually lead to different transcendentals. Since there are 2^{\aleph_0} transcendentals and 2^{\aleph_0} enumerations of the algebraic reals, perhaps all transcendentals live on diagonals. In a precise sense, they do.

Before we can state our theorem about diagonals and transcendentals, we need some definitions. Let $b(n)$ denote the n th digit of b , where b is the binary representation of a real number. We define the *diagonal number* of the sequence b_1, b_2, b_3, \dots of binary representations to be the real number whose binary representation d is obtained by the following rule: $d(n) = 0$ if $b_n(n) = 1$, and $d(n) = 1$ if $b_n(n) = 0$. (With binary representations, there is only one way to change a digit, so there is only one diagonal number associated with a sequence. To work with other representations, we would have to talk about the *diagonal numbers* of a sequence.) We say that a sequence *consists of all the binary representations of algebraic reals* if it contains all the binary representations of the algebraic reals (including both representations of the fractions $m/2^n$) and if it does not contain any representations of non-algebraic reals. Such a sequence may contain the same representation more than once. Using the above definitions, we can express the relationship between transcendentals and diagonal numbers:

Theorem 3. *A real number in the interval $(0, 1)$ is transcendental if and only if it is the diagonal number of a sequence that consists of all the binary representations of algebraic reals in $(0, 1)$.*

Proof: By Cantor's diagonal argument, the diagonal number of such a sequence is transcendental.

Now assume that t is transcendental. Let a_n be any sequence consisting of all the binary representations of algebraic reals in the interval $(0, 1)$. We will define a sequence b_n that is a permutation of the sequence a_n and that generates t as its diagonal number. -

Throughout our proof, we use the same notation to denote a real number and its binary representation. We start by finding the first a_k such that $a_k(1) \neq t(1)$. Our search is bounded by the binary representations of $1/2$ since one of these representations starts with 1 and the other starts with 0. After finding our a_k , we mark it as used and set $b_1 = a_k$. Now assume that we have found b_1, b_2, \dots, b_{n-1} . To obtain a suitable b_n , we look for the first unused a_k such that $a_k(n) \neq t(n)$. If $t(n) = 0$, then our search is bounded by the binary representations of $1/2^n + 1/2^{n+i}$ where $i = 1, \dots, n$. The representations of these numbers have a 1 in their n th place and at least one of them is unused. Similarly, if $t(n) = 1$, then our search is bounded by the binary representations of $1/2^{n+i}$ where $i = 1, \dots, n$. After finding our a_k , we mark it as used and set $b_n = a_k$.

To complete our proof, we must show that the sequence b_n is a permutation of the sequence a_n . Assume that some a_n was not used, and let a_k be the unused one with the least index. Now each a_i , for $i < k$, was used to define a b_j . Let N be greater than the indices of these b_j 's. (If $k = 1$, then there are no such b_j 's so we let N be 1.) By our definition of the sequence b_n , the only way for a_k to stay unused is for the equality $a_k(n) = t(n)$ to hold for all $n \geq N$. Hence, $t - a_k$ is rational. Since t is transcendental, a_k must also be transcendental—but this contradicts the fact that a_k is an algebraic real. Thus, the sequence b_n is a permutation of the sequence a_n .

If we apply the method in our proof to a transcendental t whose digits are computable, then we can compute a sequence of algebraic reals whose diagonal number is t . For example, we have written a computer program that uses the binary representation of $1/e$ to generate a sequence of algebraic reals whose diagonal number is $1/e$. Output from this program is given in Figure 3. The sequence generated by our program is a permutation of a sequence ω_n that consists of all the binary representations of algebraic reals in the interval $(0, 1)$. This ω_n sequence is similar to the ω_n sequence of our previous programs—the key difference is that our new sequence consists of representations rather than numbers. Since our new sequence contains both binary representations of the fractions $m/2^n$, its numbering differs from that of our previous sequence. For example, in Figure 3, both ω_1 and ω_2 are binary representations of $1/2$.

Approximation to $1/e$	Binary Representation of Algebraic Real	Associated Polynomial
.0	$\omega_2 = .1$	$2x - 1$
.01	$\omega_3 = .10 \dots$	$x^2 + x - 1$
.010	$\omega_1 = .011 \dots$	$2x - 1$
.0101	$\omega_5 = .0110 \dots$	$x^2 + 2x - 1$
.01011	$\omega_6 = .10110 \dots$	$2x^2 - 1$
.010111	$\omega_4 = .101110 \dots$	$x^2 + 2x - 2$
.0101111	$\omega_8 = .0101110 \dots$	$2x^2 + 2x - 1$
.01011110	$\omega_7 = .11000111 \dots$	$2x^2 + x - 2$
.010111100	$\omega_9 = .101010101 \dots$	$3x - 2$
.0101111000	$\omega_{10} = .0101010101 \dots$	$3x - 1$

Figure 3. Generating a sequence whose diagonal number is $1/e$.

Exercise. Construct a sequence consisting of all the binary representations of rationals in $(0, 1)$ by expanding the rationals $1/2, 1/3, 2/3, 1/4, 2/4, 3/4, \dots$ into their binary representations. Now use the algorithm in the proof of Theorem 3 to generate the first 10 members of a sequence of binary representations whose diagonal number is the irrational $\sqrt{2} - 1$. (The binary representation of $\sqrt{2} - 1$ is $0.0110101000 \dots$.)

5. CANTOR’S UNPUBLISHED PROOF. In Section 2, we saw that Cantor’s ideas can be used to write either a direct (constructive) proof or an indirect (non-constructive) proof of the existence of transcendentals. We now investigate whether Cantor knew that his ideas could produce such different proofs.

The thinking that led to Cantor's 1874 article can be found in the correspondence between Cantor and Dedekind [11, p. 187–191; 28, p. 12–16]. We start with Cantor's letter of November 29, 1873, in which he asks Dedekind the following question:

Take the collection of all positive whole numbers n and denote it by (n) ; further, imagine the collection of all positive real numbers x and denote it by (x) ; the question is simply whether (n) and (x) can be corresponded so that each individual of one collection corresponds to one and only one individual of the other.

Cantor says that at first glance it appears that no such correspondence could exist—after all, one collection is discrete and the other continuous. But then he brings out the subtlety of his question by stating that it is easy to construct a one-to-one correspondence between the collection of positive integers and the collection of rational numbers. Cantor also states that a one-to-one correspondence can be constructed between the collection of positive integers and general collections of the form $(a_{n_1, n_2, \dots, n_\nu})$, where the indices n_1, n_2, \dots, n_ν and ν are positive integers.

Dedekind replies that he is unable to answer the question, but he does give Cantor a one-to-one correspondence between the collection of algebraic numbers and the collection of positive integers. Dedekind also advises Cantor not to waste too much time on his question because it has no “particular practical interest.”

In his next letter, dated December 2nd, Cantor acknowledges Dedekind's advice but points out that his question is of interest: “It would be nice if it could be answered; for example, provided that it were answered *no*, one would have a new proof of Liouville's theorem that there are transcendental numbers.”

Cantor's letter of December 7th contains the result he is seeking. His proof starts:

Suppose that the positive numbers $\omega < 1$ can be broken up into the sequence:

$$(I) \qquad \qquad \qquad \omega_1, \omega_2, \dots, \omega_n, \dots$$

After an involved argument, Cantor obtains a contradiction.

In his next letter, dated December 9th, Cantor outlines the proof he will publish:

I show directly that if I start with a sequence

$$(I) \qquad \qquad \qquad \omega_1, \omega_2, \dots, \omega_n, \dots$$

I can determine, in *every* given interval (α, β) , a number η that is not included in (I). Hence, it follows immediately that the collection (x) cannot correspond one-to-one with the collection $(n) \dots$

Taken together, Cantor's letters of December 2nd and 7th provide an indirect proof of the existence of transcendentals. But his letter of December 9th contains his theorem on real sequences, which provides a direct construction of transcendentals.

6. WHY IS CANTOR'S ARTICLE MISINTERPRETED? Cantor's correspondence with Dedekind, which contains his indirect existence proof, was not published until 1937 [17, p. 104]. By then, other mathematicians had rediscovered the proof. Klein outlined it in 1894 [21, p. 51]. Since Klein (as far as we know) published the proof first, we will call it *Klein's proof*. In 1907, Osgood presented Klein's proof, but called it “Cantor's proof for the existence of non-algebraic numbers”

[29, p. 159–160].* In 1921, Perron presented Klein’s proof, attributed it to Cantor, and then critiqued it [30, p. 161–162]:

...Cantor’s proof for the existence of transcendental numbers has, along with all its simplicity and elegance, the great disadvantage that it is only an existence proof; it does not enable us to actually specify even a single transcendental number. Free from this disadvantage is another—in fact, the oldest—existence proof due to Liouville...

Perron’s critique is similar to the one given by Kac and Ulam (see Section 1). So this view of Cantor’s work has been around for many years. Why do some mathematicians misinterpret Cantor’s article? Undoubtedly, there are a variety of reasons. We will only discuss how these misinterpretations are encouraged by some commonly-held views about mathematics and its history.

One such view states that Cantor’s set theory was initially attacked by many mathematicians of his time. The problem with this view is that it fails to distinguish the parts of Cantor’s work that were attacked, from those that were not. For example, consider the following statement by Birkhoff and MacLane [1, p. 436–437]:

Cantor’s argument for this result [“Not every real number is algebraic”] was at first rejected by many mathematicians, since it did not exhibit any specific transcendental number.

If Cantor’s argument was rejected, then it would be reasonable to suspect that his argument is non-constructive and that this was the reason for its rejection. However, we have found no evidence indicating that Cantor’s argument was rejected. Kronecker—the mathematician most likely to reject it—had a chance to, but did not.

Cantor sent the article containing his existence proof to *Crelle’s Journal* (*Journal für die reine und angewandte Mathematik*), even though he knew that Kronecker, as one of the journal’s editors, could reject or delay the article. Previously, Kronecker had delayed the publication of an article written by Heine, one of Cantor’s colleagues. In fact, Kronecker had even tried to persuade Heine to withdraw his article [12, p. 67 and p. 308–309]. Cantor’s experience was different—his article was printed quickly. Apparently, Kronecker found it to be no worse (from his point of view) than other articles appearing in *Crelle’s Journal*.

Cantor’s article does contain revolutionary ideas, but Cantor rephrases these ideas using terminology familiar to his contemporaries. For example, he introduces the concept of a collection of reals corresponding one-to-one with the collection of positive integers, and then he provides an equivalent formulation—namely, that such a collection of reals can be written as a sequence. Also, he states the two theorems of his article in terms of sequences rather than one-to-one correspondences. Finally, Cantor incorporates a constructive idea into his article—careful reading reveals that his theorems only deal with sequences that are ordered by a “law” (see [3, p. 260; 5, p. 308] and our first Cantor quotation in Section 2). Cantor may have inserted this restriction to avoid problems with Kronecker. Kronecker required that a sequence or series be generated by an arithmetic rule—in fact, Kronecker probably objected to Heine’s article because it dealt with *arbitrary* trigonometric series [14, p. 71].

*Klein never attributed his proof to Cantor. In 1895, Klein gave an exposition of Cantor’s 1874 article in which he replaced Cantor’s old nesting method with the newer diagonal method. Klein called the resulting constructive proof “Cantor’s proof” [22, p. 49–54].

In his next article [4, 6], which was published in 1878, Cantor introduces concepts that apply to *all* sets—these concepts cannot be rephrased into the mathematical terminology of his time. Cantor begins this article by introducing the notion of a one-to-one correspondence between two arbitrary sets, finite or infinite. Next he defines the ordering that this correspondence induces—that is, what it means for one set to be of lesser, equal, or greater *power* (*cardinality*) than another. Cantor devotes most of this article to proving that for every positive integer n , the set of all n -tuples of reals can be put into one-to-one correspondence with the set of reals.

Once again, Cantor submitted his work to *Crelle's Journal*—but this time, publication was delayed. Cantor blamed Kronecker for the delay and never sent another article to *Crelle's Journal* [12, p. 69–70; 17, p. 111–113]. We have no record of why Kronecker had difficulty with Cantor's article. However, to understand this article, one must work with countably infinite and uncountably infinite sets. Kronecker, with his constructive philosophy, could never accept Cantor's reasoning.

As Cantor developed his ideas about infinity further, more mathematicians began to criticize his work. For example, Cantor presented his theory of transfinite ordinal numbers in his 1883 monograph *Foundations of a General Theory of Sets* [7, 8]. After reading Cantor's monograph, Poincaré made the following comment [13, p. 278; 27, p. 95–96]:

... these numbers in the second, and especially in the third, number-class have the appearance of being form without substance, something repugnant to the French mind.

(Cantor's first number class consists of the natural numbers. His second number class consists of the countable ordinals, which are the ordinals representing the well-orderings of the first number class. His third number class consists of the ordinals that represent the well-orderings of the second number class.)

So the popular view of the history of set theory needs to be refined. Criticism of Cantor's theory did not begin with the publication of his 1874 article. It began with his 1878 article, which contains arguments that require the use of infinite sets. Criticism increased as Cantor introduced new concepts involving the infinite.

A commonly-held view about existence proofs also needs examining—namely, the view stating that most non-constructive existence proofs are simpler than their constructive counterparts. Most non-constructive proofs are simpler. However, we must recognize those cases in which a construction yields the simplest proof. For example, Perron (see quotation above) mentions the “simplicity and elegance” of the non-constructive existence proof that we call Klein's proof. Comparing Klein's proof to Liouville's constructive proof, it is tempting to conclude that the former is simpler because of its non-constructive nature. However, Cantor's constructive approach yields an even simpler proof.

Klein's proof requires that we first prove that the set of reals is uncountable. How do we prove this? By assuming that there is a sequence that enumerates the reals, applying the diagonal method to this sequence, and obtaining a contradiction. Dissecting this proof, we find that it rests upon two facts:

- (1) If we apply the diagonal method to a sequence of reals, then we obtain a real not in the sequence.
- (2) If we assume that the reals can be enumerated by a sequence and obtain a contradiction from this assumption, then no such enumerating sequence exists.

Now if we apply the diagonal method to a sequence containing all the algebraic reals, we only need fact (1) to guarantee that we have constructed a transcendental. So Cantor's constructive approach is simpler than the non-constructive one (and it provides excellent preparation for the uncountability proof).

There may be other views about mathematics and its history that lead some mathematicians to misinterpret Cantor's article. We encourage our readers to explore this subject further.

7. CONCLUSION. We started this article with two conflicting statements about Cantor's proof of the existence of transcendentals. We have seen that these statements are talking about two different proofs of the existence of transcendentals, a constructive proof that goes back to Cantor's original articles and a non-constructive proof that appeared later.*

We advocate teaching either both proofs or just the constructive one. As mentioned in our last section, Cantor's constructive approach is simpler than the non-constructive one. Also, while discussing Cantor's approach, we can give him the credit he deserves for presenting his work so constructively.

APPENDIX. Figures 1 and 2 (see Sections 2 and 3) contain output from computer programs that use Cantor's methods. Both of these programs generate a sequence ω_n of algebraic reals. We now define this sequence precisely so that interested readers may verify our results.

As discussed in Section 2, we enumerate the polynomials with integer coefficients by their size and we use this enumeration to generate an enumeration of the algebraic reals. We only need to enumerate irreducible polynomials; but since Cantor's methods do work with sequences containing duplicates, we did not bother to check for irreducibility. (For those who are interested in testing for irreducibility, see [23, p. 431–434].) However, to simplify our calculations and avoid many duplicate roots, we decided to enumerate polynomials $p(x)$ with the following properties:

- (1) $p(x)$ has a root in the interval $(0, 1)$.
- (2) The coefficients of $p(x)$ have no common factor greater than one, and the leading coefficient is positive.
- (3) Either $p(x)$ is linear or it has no linear factors.
- (4) $p(x)$ and its derivative $p'(x)$ have no common roots.
- (5) $p(x)$ and its second derivative $p''(x)$ have no common roots.

The first condition simplifies our calculations. The second and third conditions eliminate many duplicate roots.

The last two conditions are needed by Newton's method. As Figure 1 shows (see Section 2), Cantor's 1874 nesting method generates digits very slowly. So for this method, we only used Sturm's theorem and Horner's method to approximate roots [31, p. 138–156]. But the diagonal method generates digits much faster than the nesting method. To generate a large number of these digits, we need an efficient approximation algorithm, such as Newton's method. Now to guarantee that Newton's method does converge to a root of $p(x)$, we first use Sturm's theorem and Horner's method to isolate our root to an interval where $p'(x)$ and $p''(x)$ do

*That is, appeared in print later—our study of Cantor's correspondence (see Section 5) shows that he knew both proofs.

not vanish. Since the polynomials in our enumeration do not share roots with their first or second derivatives, our root does belong to such an interval. After isolating our root, we set our initial approximation to an endpoint of this interval, and then we use Newton's method to generate better approximations (see [31, p. 174–177] for details—such as, which endpoint to choose for the initial approximation).

ACKNOWLEDGMENTS. I am very grateful to Michael Broido, Joseph Dauben, William Dunham, Stephen Maurer, Edward Sandifer, James Tattersall, and the referees for reading and commenting on drafts of this article. Their suggestions have led to numerous improvements. However, I am solely responsible for all opinions as well as any remaining errors. I am also responsible for the accuracy of all translations. Here I was aided by the French translations of Cantor's articles and letters [5, 11] as well as by Dauben's English translations [12]. My historical research was aided by Dauben's excellent biography of Cantor [12]. The photograph of Cantor on the cover is through the courtesy of the late Wilhelm Stahl and Ivor Grattan-Guinness. It shows Cantor in his middle years. Comparison with the photographs in [26, p. 530–531] suggests that it was taken between 1880 and 1894.

REFERENCES

1. G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, 4th ed., Macmillan, New York, 1977. Earlier editions: 1941, 1953, 1965.
2. E. Bishop and D. Bridges, *Constructive Analysis*, Springer-Verlag, Berlin, 1985.
3. G. Cantor, Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen, *J. Reine Angew. Math.* **77** (1874), 258–262. Reprint: [10, p. 115–118].
4. G. Cantor, Ein Beitrag zur Mannigfaltigkeitslehre, *J. Reine Angew. Math.* **84** (1878), 242–258. Reprint: [10, p. 119–133].
5. G. Cantor, Sur une propriété du système de tous les nombres algébriques réels, *Acta Math.* **2** (1883), 305–310. (French translation of [3].)
6. G. Cantor, Une contribution à la théorie des ensembles, *Acta Math.* **2** (1883), 311–328. (French translation of [4].)
7. G. Cantor, *Grundlagen einer allgemeinen Mannigfaltigkeitslehre*, Teubner, Leipzig, 1883. Reprint: [10, p. 165–208].
8. G. Cantor, Fondaments d'une théorie générale des ensembles, *Acta Math.* **2** (1883), 381–408. (French translation of the mathematical part of [7]; the philosophical part was not translated.)
9. G. Cantor, Über eine elementare Frage der Mannigfaltigkeitslehre, *Jahresbericht der Deutschen Mathematiker-Vereinigung* **1** (1891), 75–78. Reprint: [10, p. 278–280].
10. G. Cantor, *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, ed. E. Zermelo, Springer, Berlin, 1932. Reprint: Hildesheim, Olms, 1966.
11. J. Cavailles, *Philosophie mathématique*, Hermann, Paris, 1962.
12. J. W. Dauben, *Georg Cantor: His Mathematics and Philosophy of the Infinite*, Harvard University Press, Cambridge, 1978.
13. P. Dugac, *Lettres de Charles Hermite à Gösta Mittag-Leffler (1874–1883)*, Cahiers du Séminaire d'Histoire des Mathématiques **5** (1984), 49–285.
14. H. M. Edwards, Kronecker's Views on the Foundations of Mathematics, *The History of Modern Mathematics, Vol. 1*, (edited by D. E. Rowe and J. McCleary), Academic Press, Boston, 1989, 67–77.
15. A. Fraenkel, Georg Cantor, *Jahresbericht der Deutschen Mathematiker-Vereinigung* **39** (1930), 189–266.
16. M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, San Francisco, 1979.
17. I. Grattan-Guinness, The rediscovery of the Cantor-Dedekind correspondence, *Jahresbericht der Deutschen Mathematiker-Vereinigung* **76** (1974), 104–139.
18. R. Gray, The computational complexity of Cantor's transcendentals, (in preparation).
19. I. N. Herstein and I. Kaplansky, *Matters Mathematical*, Harper & Row, New York, 1974.
20. M. Kac and S. M. Ulam, *Mathematics and Logic*, Frederick A. Praeger, New York, 1968.
21. F. Klein, *Lectures on Mathematics*, Macmillan, New York, 1894.
22. F. Klein, *Vorträge über Ausgewählte Fragen der Elementargeometrie*, Teubner, Leipzig, 1895. Translated by W. W. Beman and D. E. Smith as *Famous Problems of Elementary Geometry*, Ginn, Boston, 1897; reprinted by Chelsea Publishing Company, New York, 1955. Page references are to English translation.

23. D. Knuth, *The Art of Computer Programming Vol. 2*, 2nd ed., Addison-Wesley, Reading, Mass., 1981.
24. J. Liouville, Remarques sur des classes très-étendues de quantités dont la valeur n'est ni algébrique, ni même réductible à des irrationnelles algébriques, *Comp. Rend. Acad. Sci. Paris* 18 (1844), 883–885.
25. J. Liouville, Sur des classes très-étendues de quantités dont la valeur n'est ni algébrique, ni même réductible à des irrationnelles algébriques, *J. Math. Pures Appl.* 16 (1851), 133–142.
26. H. Meschkowski and W. Nilson, eds., *Georg Cantor Briefe*, Springer-Verlag, Berlin, 1991.
27. G. H. Moore, Towards A History of Cantor's Continuum Problem, *The History of Modern Mathematics, Vol. 1*, (edited by D. E. Rowe and J. McCleary), Academic Press, Boston, 1989, 79–121.
28. E. Noether and J. Cavaillès, eds., *Briefwechsel Cantor-Dedekind*, Hermann, Paris, 1937.
29. W. F. Osgood, *Lehrbuch der Funktionentheorie*, Teubner, Leipzig, 1907. Reprint of 1928 edition: Chelsea, New York, 1965. (In reprint, Cantor's work is mentioned on p. 203–204.)
30. O. Perron, *Irrationalzahlen*, de Gruyter, Berlin, 1921. Reprint of 1929 edition: Chelsea, New York, 1951. (In reprint, Cantor's work is mentioned on p. 174.)
31. J. V. Uspensky, *Theory of Equations*, McGraw-Hill, New York, 1948.

5 Joseph Road
Framingham, MA 01701

Added in Proof. The quotation of Poincaré in Section 6 is out of context. It comes from a letter explaining why French mathematicians may not appreciate a proposed translation of Cantor's 1883 monograph—a monograph that Poincaré twice calls “beautiful” [13, p. 278]. Poincaré warns that, unless they are given concrete examples, mathematicians unfamiliar with Cantor's previous work may find his ordinals to be “form without substance.”

A quotation that accurately reflects some of the criticism of the time can be found in an 1883 letter from Hermite to Mittag-Leffler [13, p. 209; 27, p. 96]:

The impression that Cantor's memoirs makes on us is distressing. Reading them seems, to all of us, to be a genuine torture While recognizing that he has opened up a new field of research, none of us is tempted to pursue it. For us it has been impossible to find, among the results that can be understood, a single one having *current interest*. The correspondence between the points of a line and a surface leaves us absolutely indifferent and we think that this result, as long as no one has deduced anything from it, stems from such arbitrary methods that the author would have done better to withhold it and wait.

Hermite's “us” includes Appell, who read a draft of the letter, and probably Picard, who was also critical of Cantor's recent work [13, p. 212]. The criticism at the end of the quotation is directed at the major result of Cantor's 1878 article.

The 500th Anniversary of the Sharing Problem (The Oldest Problem in the Theory of Probability)

Milton Sobel and Krzysztof Frankowski

1. INTRODUCTION. The problem of sharing (also known as the Problem of Points, Problème des Parties, Teilungsproblem) is the oldest known problem in probability theory according to K. Jordan [2] having been described in a book of the famous mathematician Lucas de Burgo Pacioli published in 1494 [3], about 500 years ago, in the ancient Italian language. It occupied the attention of many illustrious mathematicians including Fermat, Pascal, Huygens, Ja. Bernoulli, Montmort, De Moivre, Laplace, Lagrange, Tartaglia and Pacioli, to name a few. The 2 and 3-player problems with unequal parameters and unequal cell probabilities were solved long ago by the people listed above (cf. K. Jordan [2] and Todhunter [6] for details) but no unified treatment for the general case for equal or unequal cell probabilities to our knowledge has ever been published. In this paper the recently developed Dirichlet integrals [4], [5] are used to give a unified treatment of the general case with k players and equal or unequal cell probabilities. In addition we extend the problem by assuming that the contest might be continued at some future time and given the past results as initial conditions we derive results:

- i) for the expected waiting time (first moment) to complete the tournament and also higher moments,
- ii) for the probability of each possible vector-ranking of the players at the completion of the tournament,
- iii) for four different disciplines used in completing the tournament,
- iv) allowing the possibility of a quota-free cell, so that the sum of the cell probabilities for the competitive players may be less than unity, and
- v) allowing ties which clearly affect the expected waiting time, but without giving any fractional credit for additional games won.

The problem is quite simple to describe. A tournament with k players consist of games with all k players involved in each game. Each game produces 1 winner and the games are carried out sequentially until one of the players wins n games. At some point when the i th player has won $w_i < n$ games ($i = 1, 2, \dots, k$), or needs $n - w_i = r_i > 0$ games to win the tournament, the contest is suddenly discontinued. How should the prize money be shared if we wish to distribute it in proportion to each player's conditional probability of winning the tournament, assuming that the contest were continued? The probability $p_i > 0$ that the i th player wins in a single game ($i = 1, 2, \dots, k$) is given and the games are independent of each other.

In the original formulation there is no chance of a tie in any game. We later introduce an additional quota-free cell with cell probability $p_0 = 1 - \sum_{i=1}^k p_i < 1$,

which can also be regarded as the probability of a tie; nobody gets any fractional credit for a tied game, but it does add to the expected waiting time for completion of the tournament. On the other hand the expression “sink” could indicate that we are singling out the cell of a particular one of the k players and waiting until he reaches a fixed number of wins and hence it may not refer to an additional cell. If r_i -values are regarded as quotas, then only the cell associated with p_0 is quota free. If there is only 1 cell (with probability p_0) that is used for stopping, we call it a counting cell. The quantity b is used in the tables in [4] and [5] and is usually equal to k or $k - 1$; $b = k$ if there are k competitive players and we are not using any player as a sink, but if we use one player as a sink for a computation then $b = k - 1$ in that computation, regardless of whether a quota-free cell is present or not. Hence the use of the symbol b in a formula (instead of k) gives us more flexibility.

2. BACKGROUND TOOLS AND THEIR NOTATION. The Dirichlet integrals consist of two types: Type 1 (includes the I , the J and the II) and Type 2 (includes the C , the D and the CD); the integrals of Type 2 constitute the main tool used in this paper. These integrals are defined, studied, tabulated with many illustrations of usage in [5] and we repeat them briefly.

For any positive integer b and m with vectors $\vec{a} = (a_1, \dots, a_b)$ (real) and $\vec{r} = (r_1, \dots, r_b)$ (natural) we define

$$C_{\vec{a}}^{(b)}(\vec{r}, m) = \frac{\Gamma(m + R)}{\Gamma(m)\prod_{i=1}^b \Gamma(r_i)} \int_0^{a_1} \cdots \int_0^{a_b} \frac{\prod_{i=1}^b x_i^{r_i-1} dx_i}{(1 + \sum_{i=1}^b x_i)^{m+R}}, \quad (2.1)$$

where $R = r_1 + \dots + r_b$, $a_i = p_i/p_0$ ($i = 1, 2, \dots, b$), p_i are the cell probabilities and $p_0 = 1 - \sum_{i=1}^b p_i < 1$.

For the probability interpretation of (2.1) we consider a multinomial setting with b blue cells, where the i th cell has cell probability p_i and quota r_i . In addition we have a counting cell with cell probability p_0 . We sample from this multinomial until the counting cell reaches frequency $m \geq 1$ for the first time, referring to this as “at stopping time” or *ast*. Then (2.1) is the probability that for each of the b blue cells, the i th cell has its quota of at least r_i *ast*.

The dual probability that for each of the b cells, the i th cell has less than r_i ($i = 1, \dots, b$) *ast* is given by:

$$D_{\vec{a}}^{(b)}(\vec{r}, m) = \frac{\Gamma(m + R)}{\Gamma(m)\prod_{i=1}^b \Gamma(r_i)} \int_{a_1}^{\infty} \cdots \int_{a_b}^{\infty} \frac{\prod_{i=1}^b x_i^{r_i-1} dx_i}{(1 + \sum_{i=1}^b x_i)^{m+R}}, \quad (2.2)$$

with the same notation as in (2.1). We need also the mixed integral, where $0 \leq c \leq b$ integrals are from 0 to a_i or of the C-type while the remaining $b - c$ integrals are from a_i to ∞ or of the D-type, namely

$$CD_{\vec{a}}^{(c, b-c)}(\vec{r}, m) = \frac{\Gamma(m + R)}{\Gamma(m)\prod_{i=1}^b \Gamma(r_i)} \int_0^{a_1} \cdots \int_0^{a_c} \int_{a_{c+1}}^{\infty} \cdots \int_{a_b}^{\infty} \frac{\prod_{i=1}^b x_i^{r_i-1} dx_i}{(1 + \sum_{i=1}^b x_i)^{m+R}}. \quad (2.3)$$

The probability interpretation is given in the same multinomial setting, with the obvious difference that for each of the c cells the i th cell ($i = 1, \dots, c$) has its quota of at least r_i and for the rest of the $b - c$ cells the i th cell ($i = c + 1, \dots, b$) has less than r_i . For $b = 0$ we define $C^{(0)} = D^{(0)} = CD^{(0,0)} = 1$.

In the homogeneous case we have all $p_i = p (i = 1, 2, \dots, b)$, $p_0 = 1 - bp$ and we use the common scalar $a = p/p_0$ instead of \vec{a} . Whenever we use more than one r (or s) value in a Dirichlet function we separate them from the last argument with a semicolon.

For the proof in Section 3 we need a slight generalization of (2.1). For the present $\vec{r} = (r_1, r_2, \dots, r_b)$ and $\vec{a} = (a_1, a_2, \dots, a_b)$ are each b vectors. Let X_i denote the frequency in the i th blue cell ($i = 1, 2, \dots, b$) and $R = r_1 + r_2 + \dots + r_b$ as before. Let $P = C_{\vec{a}}^{(b,j)}(\vec{r}; m)$ denote the probability, when the counting cell reaches m (for the first time), that the joint event $X_i = r_i$ ($i = 1, 2, \dots, j$) and $X_\alpha \geq r_\alpha$ ($\alpha = j + 1, j + 2, \dots, b$) for given j with $0 \leq j < b$ has occurred. Then we have

$$P = \frac{\Gamma(m + R)}{\Gamma(m)} \frac{\prod_{i=1}^j (a_i^{r_i}/r_i)}{\prod_{i=1}^b \Gamma(r_i)} \int_0^{a_{j+1}} \dots \int_0^{a_b} \frac{\prod_{\alpha=j+1}^b x_\alpha^{r_\alpha-1} dx_\alpha}{(1 + \sum_{i=1}^j a_i + \sum_{\alpha=j+1}^b x_\alpha)^{m+R}} \quad (2.4)$$

Clearly the j cells with equality are taken to be the first j cells only for notational convenience. For $j = 0$ we drop the second superscript as in (2.1) and for $j = b$ we get multinomial probabilities as in (2.26) of [5]. The same result (2.4) holds for $D_{\vec{a}}^{(b,j)}(\vec{r}; m)$, where the event changes to $X_\alpha < r_\alpha$ ($\alpha = j + 1, \dots, b$), except that all the limits of integration are from a_α to ∞ .

We also use in Section 4 (and elsewhere) the multiple multinomial sum expression for the Dirichlet D integral as follows:

$$D_{\vec{a}}^{(b)}(\vec{r}, m) = \frac{1}{(1 + \sum_{i=1}^b a_i)^m} \sum_{x_1 < r_1} \dots \sum_{x_b < r_b} \left[\begin{matrix} m - 1 + \sum_{\alpha=1}^b x_\alpha \\ m - 1, x_1, \dots, x_b \end{matrix} \right] \times \prod_{i=1}^b \left(\frac{a_i}{1 + \sum_{\alpha=1}^b a_\alpha} \right)^{x_i}, \quad (2.5)$$

where the square bracket denotes the multinomial coefficient; for the case of common a and r , this result appears as (2.15) in [5].

3. PROBABILITY RESULTS FOR THE SHARING PROBLEM. The D and C Dirichlet integrals provide an immediate unified solution to the generalized problem with k player with equal or unequal cell probabilities. Assuming the tournament is continued, the first player to reach n wins is clearly the winner of the tournament; if the contestants continue beyond that, we say that the last player to reach n wins is the loser (or the k th rank winner) of the tournament, i.e., he is the last player to reach his quota. If we order all k players in order that they reach their quota, we call it a vector-ranking of the players and we are also interested in the probability of any particular vector-ranking.

Let $\vec{a}_i = (p_1/p_i, \dots, p_{i-1}/p_i, p_{i+1}/p_i, \dots, p_k/p_i)$ and $\vec{r}_i = (r_1, \dots, r_{i-1}, r_{i+1}, \dots, r_k)$, both with $b = k - 1$ components. [Here we are singling out the i th player and getting ready to use him as a sink.] From the probability interpretation of the D integral the probability W_i that player i reaches r_i games won before any player j reaches r_j games won ($j \neq i$) is (exactly)

$$W_i = D_{\vec{a}_i}^{(b)}(\vec{r}_i; r_i) \quad (3.1)$$

and this interpretation also gives us the identity

$$\sum_{i=1}^k W_i = \sum_{i=1}^k D_{a_i}^{(b)}(\vec{r}_i; r_i) = 1, \quad (3.2)$$

since the probability of a tie, namely p_0 , is less than 1 and hence one of the k players has to win the tournament. Both of the above results hold regardless of whether quota-free cells are present or not.

Similarly the probability that player i is the loser of the tournament is

$$L_i = C_{a_i}^{(b)}(\vec{r}_i; r_i) \quad (3.3)$$

and

$$\sum_{i=1}^k L_i = \sum_{i=1}^k C_{a_i}^{(b)}(\vec{r}_i; r_i) = 1. \quad (3.4)$$

To illustrate (3.2) we repeat one of the seventeen calculations from the short but classical table of Huygens [1], republished in [2]. In this example we have 3 players, with common probability $p = 1/3$ and quotas $r_1 = 2, r_2 = 3, r_3 = 4$. From the definition (2.2) we have, integrating by parts,

$$\begin{aligned} W_1 &= D_1^{(2)}(3, 4; 2) = \frac{\Gamma(9)}{\Gamma(2)\Gamma(3)\Gamma(4)} \int_1^\infty \int_1^\infty \frac{x^2 y^3 dx dy}{(1+x+y)^9} \\ &= \frac{1}{8} D_{1/2}^{(1)}(3, 5) + D_1^{(2)}(3, 2) \approx .6186557, \end{aligned} \quad (3.5)$$

$$\begin{aligned} W_2 &= D_1^{(2)}(2, 4; 3) = \frac{\Gamma(9)}{\Gamma(2)\Gamma(3)\Gamma(4)} \int_1^\infty \int_1^\infty \frac{xy^3 dx dy}{(1+x+y)^9} \\ &= \frac{3}{16} D_{1/2}^{(1)}(4, 4) + \frac{1}{8} D_{1/2}^{(1)}(4, 3) \approx .2674897, \end{aligned} \quad (3.6)$$

$$\begin{aligned} W_3 &= D_1^{(2)}(2, 3; 4) = \frac{\Gamma(9)}{\Gamma(2)\Gamma(3)\Gamma(4)} \int_1^\infty \int_1^\infty \frac{xy^2 dx dy}{(1+x+y)^9} \\ &= \frac{1}{8} D_{1/2}^{(1)}(3, 5) + \frac{1}{16} D_{1/2}^{(1)}(3, 4) \approx .1138546. \end{aligned} \quad (3.7)$$

Each of the answers in (3.5), (3.6), (3.7) were obtained by reducing the D -functions to a single common r argument and then using the 8-place table in [5]. One could also finish integration and get exact values $451/729$, $195/729$, and $83/729$, but since each of the given answers is correct to 8 decimals there is no need for that. Note that these answers sum to one, i.e., (3.2) is satisfied.

The three C -integrals with exactly the same arguments give us, respectively, the probability that player j will lose the tournament. Since the limits in this case are both from 0 to 1 there is a bit more calculation and the tables become more useful. The exact results are respectively

$$\begin{aligned} L_1 &= \frac{451}{729} - \frac{16}{32} \approx .1186557, & L_2 &= \frac{195}{729} + \frac{1}{32} \approx .2987397, \\ L_3 &= \frac{83}{729} + \frac{15}{32} \approx .5826046. \end{aligned} \quad (3.8)$$

Since these also sum to one, (3.4) is satisfied.

Let $P(r_1, r_2, r_3)$ denote the joint probability that the player, who needed r_i more wins, would have rank i if the contest were completed ($i = 1, 2, 3$). For the case of 3 players we can use the probability interpretation of (2.3) and in our

example this probability is

$$P(r_1, r_2, r_3) = CD_1^{(1,1)}(2, 4; 3) = \frac{\Gamma(9)}{\Gamma(2)\Gamma(3)\Gamma(4)} \int_0^1 \int_1^\infty \frac{xy^3 dydx}{(1+x+y)^9} \approx .3887603, \quad (3.9)$$

and for the remaining five vector rankings we obtain

$$\begin{aligned} P(r_1, r_3, r_2) &= CD_1^{(1,1)}(2, 3; 4) \approx .2298954, \\ P(r_2, r_1, r_3) &= CD_1^{(1,1)}(3, 4; 2) \approx .1938443, \\ P(r_2, r_3, r_1) &= CD_1^{(1,1)}(3, 2; 4) \approx .0736454, \\ P(r_3, r_1, r_2) &= CD_1^{(1,1)}(4, 3; 2) \approx .0688443, \\ P(r_3, r_2, r_1) &= CD_1^{(1,1)}(4, 2; 3) \approx .0450103, \end{aligned} \quad (3.10)$$

all of which also sum to one. By combining the appropriate pairs we can also obtain the vector results for any one player. Thus if $P_1(R_j)$ denotes the probability that player 1 ends up with rank j ($j = 1, 2, 3$) we have from the above results for player 1

$$\{P_1(R_1), P_1(R_2), P_1(R_3)\} = \left\{ \frac{902}{1458}, \frac{383}{1458}, \frac{173}{1458} \right\}, \quad (3.11)$$

which also add to one.

In the case of 4 or more players the probability of any particular vector-ranking is not nearly as simple as in the case of 3 players. For notational convenience below we replace (r_1, r_2, r_3, r_4) by (f, s, t, r) respectively, and use Δ to denote $p_2 + p_3 + p_4$.

Theorem. *The probability $P(\vec{v}^0)$ of the vector ranking $\vec{v}^0 = (1, 2, 3, 4)$ of 4 players with quotas f, s, t , and r and cell probabilities p_i ($i = 1, 2, 3, 4$) with $p_1 + \Delta \leq 1$ is given by the double summation with tr terms*

$$\begin{aligned} P(\vec{v}^0) &= \sum_{i=0}^{t-1} \sum_{j=0}^{r-1} \left[\begin{matrix} s-1+i+j \\ s-1, i, j \end{matrix} \right] \left(\frac{p_2}{\Delta} \right)^s \left(\frac{p_3}{\Delta} \right)^i \left(\frac{p_4}{\Delta} \right)^j \\ &\quad \times C_{p_1/\Delta}^{(1)}(f, s+i+j) D_{p_4/p_3}^{(1)}(r-j, t-i). \end{aligned} \quad (3.12)$$

For common p the five ratios in (3.12) are all $1/3$ or 1 and the result does not depend on p , only on the quotas. If the quotas are also common then the value is $1/4! = 1/24$ for any vector \vec{v} .

Sketch of the Proof: We simply condition on the number of games won by the third and the fourth player at the time when the second player reaches his quota s ; call these i and j respectively. Then we can use the result (2.4) with $j = 2$ and $b = 3$ with player 1 having attained his quota f , player 2 just reaching s and player 3 and 4 temporarily at i and j wins respectively. In the ensuing games player 3 has to reach $t - i$ before player 4 reaches $r - j$. Thus we have

$$P(\vec{v}^0) = \sum_{i=0}^{t-1} \sum_{j=0}^{r-1} C_{(p_1/p_2, p_3/p_2, p_4/p_2)}^{(3,2)}(f, i, j; s) D_{p_4/p_3}^{(1)}(r-j, t-i). \quad (3.13)$$

Using only the integral expression for C from (2.4) into (3.13) we factor out from the integrand the quantity $1 + p_3/p_2 + p_4/p_2$ and with an obvious change of variable we obtain the result (3.12).

We will illustrate this theorem using the case of the four players with quotas $r_i = i$, ($i = 1, 2, 3, 4$) and common $p = 1/4$. From (3.1) using (2.2), we obtain

$$W_1 = D_1^{(3)}(2, 3, 4; 1) \approx .6921387, \quad W_2 = D_1^{(3)}(1, 3, 4; 2) \approx .2150879, \quad (3.14)$$

$$W_3 = D_1^{(3)}(1, 2, 4; 3) \approx .0700684, \quad W_4 = D_1^{(3)}(1, 2, 3; 4) \approx .0227051. \quad (3.15)$$

Using (3.12) we obtain the desired numerical probabilities for the $4! = 24$ vector rankings v^0 which we arrange in the following table

Vector-ranking probabilities for $\vec{r} = (1, 2, 3, 4)$; common $p = 1/4$

v^0	$P(v^0)$	v^0	$P(v^0)$	v^0	$P(v^0)$	v^0	$P(v^0)$
1234	.2508208	2134	.0955250	3124	.0292612	4123	.0070258
1243	.1527470	2143	.0573295	3142	.0111104	4132	.0045392
1324	.1409879	2314	.0334248	3214	.0187104	4213	.0049145
1342	.0564335	2341	.0089897	3241	.0048848	4231	.0022199
1423	.0546842	2413	.0135274	3412	.0035737	4312	.0023510
1432	.0364654	2431	.0062916	3421	.0025278	4321	.0016547
sums	.6921388		.2150880		.0700683		.0227051

As the last row indicates, summing up the columns gives us the values for W_i , $i = 1, 2, 3, 4$ again as given in (3.14) and (3.15), except for rounding errors.

The basic difficulty in developing the expression (3.13) for the probability of vector rankings was in getting from 3 to 4 players. The method used for 4 players can be generalized in a straightforward manner to any number of players, but not without increasing the number of summations, the complexity of the analysis and the space needed to write down $k!$ numerical solutions. Thus for 5 players in the analogue of (3.12) and (3.13) we would have 3 summations on the product of $C^{(4,3)}$ and $CD^{(1,1)}$, i.e., we first use (2.4) with $j = 3$, $b = 4$ with player 1 having attained his quota r_1 , player 2 just reaching r_2 and players 3, 4 and 5 temporarily at h , i and j wins respectively. In the ensuing games when player 4 reaches $r_4 - i$ wins, player 3 must already have reached $r_3 - h$ wins and player 5 must not yet have reached $r_5 - i$ wins; this is easily expressed by a $CD^{(1,1)}$ function. Hence we only need 1 extra summation to write the analogue of (3.13) for 5 players, namely for the vector ranking $\vec{v}^0 = (1, 2, 3, 4, 5)$ of 5 players with quotas r_1, r_2, r_3, r_4 and r_5 and cell probabilities p_i ($i = 1, 2, 3, 4, 5$). Letting $\Delta = p_2 + p_3 + p_4 + p_5$ with $p_1 + \Delta \leq 1$, we have

$$\begin{aligned}
p(\vec{v}^0) &= \sum_{h=0}^{r_3-1} \sum_{i=0}^{r_4-1} \sum_{j=0}^{r_5-1} C_{(p_1/p_2, p_3/p_2, p_4/p_2, p_5/p_2)}^{(4,3)}(r_1, h, i, j; r_2) \\
&\quad \times CD_{(p_3/p_4, p_5/p_4)}^{(1,1)}(r_3 - h, r_5 - j; r_4 - i) \\
&= \sum_{h=0}^{r_3-1} \sum_{i=0}^{r_4-1} \sum_{j=0}^{r_5-1} \left[\begin{matrix} r_2 - 1 + h + i + j \\ r_2 - 1, h, i, j \end{matrix} \right] \left(\frac{p_2}{\Delta} \right)^{r_2} \left(\frac{p_3}{\Delta} \right)^h \left(\frac{p_4}{\Delta} \right)^i \left(\frac{p_5}{\Delta} \right)^j \\
&\quad \times C_{p_1/\Delta}^{(1)}(r_1, r_2 + h + i + j) CD_{(p_3/p_4, p_5/p_4)}^{(1,1)}(r_3 - h, r_5 - j; r_4 - i),
\end{aligned} \tag{3.16}$$

which was obtained by a method analogous to that used for (3.12).

4. EXPECTATION (WAITING TIME) RESULTS FOR THE SHARING PROBLEM. We refer to the time point when the contest is stopped and player i still needs r_i games to win the tournament ($i = 1, 2, \dots, k$) as the ‘time of abortion’. A related interesting problem (that appears *not* to have been considered) is to find

the expectation and/or distribution of the expected additional number of games (after abortion) needed to complete the tournament. We also consider continuing beyond that until a complete ranking or some other preassigned partial ranking is established. For the complete ranking we consider 4 different disciplines called D_0 , D_1 , D_2 and D_3 . Under discipline D_1 early winners stop playing as soon as they reach their quota (or r -value). If player i drops out we divide all the other cell probabilities by $1 - p_i$, ($i = 1, 2, \dots, k$), so that the sum is still one and the ratios are unchanged. Under discipline D_2 we stop the entire contest after the player with the rank $k - 1$ reaches his quota, so that the last one is determined. Under discipline D_3 (resp., D_0) we include both (resp., neither) of these two features. Our principal interest for more than one player reaching their quotas is in discipline D_0 where we continue with *all* players until a desired goal is reached. Again we do not assume equal p -values. As in [5] the symbol $\mu_i^{[\gamma]}$ will denote the γ^{th} ascending factorial moment of the additional waiting time needed under the discipline D_i ($\gamma = 0, 1, \dots; i = 0, 1, 2, 3$) and $x^{[\gamma]}$ will denote the product $x(x + 1) \dots (x + \gamma - 1)$. If we wait only for the first winner of the tournament then we omit the subscript indicating the discipline. We first carry out the solution for $k = 3$ players and then write the result for $k (= b + 1)$ players since the argument is essentially the same. Our development is based partly on the expression in (2.5) of the D -function as a (multiple) multinomial sum.

For $k = 3$ players the probability that player j has $< r_j$ (say, α_j) won games when player i wins the tournament with r_i games won ($i = 1, 2, 3$) is obtained by summing (over $0 \leq \alpha_j < r_j, j \neq i$) the expression for $i = 1$

$$\begin{bmatrix} r_1 - 1 + \alpha_2 + \alpha_3 \\ r_1 - 1, \alpha_2, \alpha_3 \end{bmatrix} p_1^{r_1} p_2^{\alpha_2} p_3^{\alpha_3} + 2 \text{ similar terms for } i = 2 \text{ and } 3. \quad (4.1)$$

Multiplying these 3 terms by $(r_1 + \alpha_2 + \alpha_3)^{[\gamma]}$, $(r_2 + \alpha_2 + \alpha_3)^{[\gamma]}$ and $(r_3 + \alpha_2 + \alpha_3)^{[\gamma]}$ respectively and summing over α 's gives us the desired γ^{th} ascending factorial moment until the first (cf. argument 1) tournament winner as

$$\begin{aligned} \mu^{[\gamma]}(1) &= \frac{r_1^{[\gamma]}}{p_1^\gamma} \sum_{\substack{\alpha_2 < r_2 \\ \alpha_3 < r_3}} \begin{bmatrix} r_1 + \gamma - 1 + \alpha_2 + \alpha_3 \\ r_1 + \gamma - 1, \alpha_2 + \alpha_3 \end{bmatrix} p_1^{r_1 + \gamma} p_2^{\alpha_2} p_3^{\alpha_3} + 2 \text{ similar terms} \\ &= \frac{r_1^{[\gamma]}}{p_1^\gamma} D_{\frac{p_2}{p_1}, \frac{p_3}{p_1}}^{(2)}(r_2, r_3; r_1 + \gamma) + 2 \text{ similar terms.} \end{aligned} \quad (4.2)$$

Using the notation \vec{a}_i and \vec{r}_i of Section 3 above (see (3.1)) the general result for $k = b + 1$ players can be written succinctly as

$$\mu^{[\gamma]}(1) = \sum_{i=1}^k \frac{\Gamma(r_i + \gamma)}{\Gamma(r_i) p_i^\gamma} D_{\vec{a}_i}^{(b)}(\vec{r}_i; r_i + \gamma); \quad (4.3)$$

this should be compared with equation (5.8) of [5]. The same formula (4.3) holds even if there is a dealer connected with the contest who has a p -value for each game but no quota, provided he is not included in the value of k which represents the number of active competitors. In other words the same result holds if $p_1 + p_2 + \dots + p_k < 1$ and we have a quota-free cell present. The use of b for $k - 1$ in (4.3) is to *avoid* any confusion since b is the parameter used in the tables in [4] and [5] and is also used in the definitions in Section 2 above.

In a similar manner we can obtain for any given s ($1 \leq s \leq k$) the additional expected number of games until the player with rank s meets his quota, i.e., until

the first time that any s of the k players meet their quota. For fixed i and s we define $\vec{a}_{i;\beta}$ as a permutation of \vec{a}_i where the parameter β specifies combinations of $s - 1$ of the components of \vec{a}_i so that β runs from 1 to $\binom{k-1}{s-1}$. This result needs a double summation, i specifies the player of rank s ($1 \leq i \leq k$) and β selects $s - 1$ out of $k - 1$ for the first $s - 1$ tournament winners. The result under discipline D_0 is

$$\mu_0^{[\gamma]}(s) = \sum_{i=1}^k \frac{\Gamma(r_i + \gamma)}{\Gamma(r_i) p_i^\gamma} \sum_{\beta=1}^{\binom{k-1}{s-1}} CD_{\vec{a}_{i;\beta}}^{(s-1, k-s)}(\vec{r}_{i;\beta}; r_i + \gamma), \quad (4.4)$$

where $\vec{r}_{i;\beta}$ selects the same subset from \vec{r}_i that $\vec{a}_{i;\beta}$ does from \vec{a}_i . Note that the sum of the superscripts is again $k - 1 = b$.

For $s = 1$ the result (4.4) reduces to the previous result (4.3) and for $s = b$ we get the same as (4.3) with D replaced by C . This is the expectation under discipline D_0 if we wait for every player to reach his quota.

Illustrations. For the example of Huygens with $k = 3$, common $p = 1/3$ and $\vec{r} = \{2, 3, 4\}$ we obtain from (4.3) for the first winner ($s = 1$)

$$\mu(1) = 6D_1^{(2)}(3, 4; 3) + 9D_1^{(2)}(2, 4; 4) + 12D_1^{(2)}(2, 3, 5) \approx 4.3909465, \quad (4.5)$$

$$\mu^{[2]}(1) = 54D_1^{(2)}(3, 4; 4) + 108D_1^{(2)}(2, 4; 5) + 180D_1^{(2)}(2, 3; 6) \approx 25.7695168, \quad (4.6)$$

$$\sigma^2(1) = \mu^{[2]}(1) - \mu(1)(1 + \mu(1)) \approx 2.0981590. \quad (4.7)$$

In another example with $k = 3$ players having common $p = 1/3$ and $\vec{r} = (2, 1, 1)$, we wait for 2 players to reach their quota, i.e., we take $s = 2$ and use discipline D_0 . Then from (4.4) for $\gamma = 1$ because of symmetry the 6 values reduce to 3 and we obtain

$$\begin{aligned} \mu_0(2) &= 12CD_1^{(1,1)}(1, 1; 3) + 6CD_1^{(1,1)}(1, 2; 2) + 6CD_1^{(1,1)}(2, 1, 2) \\ &= 12[D_1^{(1)}(1; 3) - D_1^{(2)}(1; 3)] + 6D_1^{(1)}(2; 2) + 6D_1^{(1)}(1; 2) - 12D_1^{(2)}(1, 2; 2) \\ &= 12\left(\frac{1}{8} - \frac{1}{27}\right) + 3 + \frac{6}{4} - 12\left(\frac{5}{27}\right) = \frac{10}{3}. \end{aligned} \quad (4.8)$$

In the same illustration for $s = 1$ we obtain from (4.3) for $\gamma = 1$

$$\mu(1) = 6D_1^{(2)}(1, 1; 3) + 6D_1^{(2)}(1, 2; 2) = 2/9 + 10/9 = 4/3. \quad (4.9)$$

The variance for this example from (4.3) is

$$\sigma^2 = \mu^{[2]}(1) - \mu(1)(1 + \mu(1)) = 54D_1^{(2)}(1, 1; 4) + 36D_1^{(2)}(1, 2; 3) - \left(\frac{4}{3}\right)\left(\frac{7}{3}\right) = \frac{2}{9}. \quad (4.10)$$

The results in (4.8), (4.9) and (4.10) can also be checked by a complete enumeration, which we omit.

In the next section we consider the effect of having early winners withdraw from the contest when $s \geq 2$, i.e., when the discipline is D_1 rather than D_0 and we wait for at least 2 players to reach their quota.

5. EARLY WINNERS WITHDRAW FROM THE TOURNAMENT (DISCIPLINE D_1). It should be noted that the probabilities of vector ranking are not affected by early withdrawal of winners. Actually we carry out the expected waiting time calculation only for $s = 2$ and $\gamma = 1$ but k remains arbitrary; for the case $k = 3$

this already gives us the expected minimum number of games needed to determine the ranking of all 3 players.

Let $\vec{r}_{i,j}$ and $\vec{a}_{i,j}$ denote vectors as before but with both the i and j components removed and using the common denominator p_j in the $\vec{a}_{i,j}$ vector. Then the required expectation $\mu_1(2)$ under discipline D_1 for $s = 2$ and $\gamma = 1$ is given by

$$\begin{aligned} \mu_1(2) = & \sum_{i \neq j} (r_i + r_j) CD_{\vec{a}_{i,j}}^{(1,k-2)}(r_i, \vec{r}_{i,j}; r_j) \\ & + \sum_{i \neq j} \frac{r_j}{p_j} \sum_{\alpha \neq i,j} p_\alpha CD_{\vec{a}_{i,j}}^{(1,k-2)}(r_i, \vec{r}_{i,j}^\alpha; r_j + 1), \end{aligned} \quad (5.1)$$

where $\vec{r}_{i,j}^\alpha = \vec{r}_{i,j}$ with r_α replaced by $r_\alpha - 1$ for some $\alpha \neq i, j$ associated with a D -type integral and each outside summation has $k(k-1)$ terms with i denoting the first winner and j the second winner, i.e., the pairs (i, j) are ordered. Here α sums over all $k-2$ indices associated with D -type integrals and if in any term $r_\alpha = 1$ then $r_\alpha - 1 = 0$ and that term vanishes. In particular for common $r = 1$ and common $p = 1/3$ from (5.1) we get $\mu_1(2) = 6(2)(1/6) + 6(0) = 2$, which is clearly the correct answer. Although (5.1) contains our desired result for $s = 2$, $\gamma = 1$ and arbitrary k , the extension of (5.1) to any s and any γ has yet to be done.

Proof of (5.1): Assume that r_1 corresponds to the first tournament winner and r_2 to the second; all remaining terms can be obtained by simply permuting the arguments. The basic idea is already present for $k = 3$ and we restrict our proof to this case. Consider the expression TS given by

$$\begin{aligned} TS = & \sum_{\substack{\alpha_2 < r_2 \\ \alpha_3 < r_3}} \left[\begin{matrix} r_1 - 1 + \alpha_2 + \alpha_3 \\ r_1 - 1, \alpha_2, \alpha_3 \end{matrix} \right] p_1^{r_1} p_2^{\alpha_2} p_3^{\alpha_3} \sum_{\beta_3 < r_3 - \alpha_3} Q \left[\begin{matrix} r_2 - \alpha_2 - 1 + \beta_3 \\ r_2 - \alpha_2 - 1, \beta_3 \end{matrix} \right] \\ & \times \left(\frac{p_2}{p_2 + p_3} \right)^{r_2 - \alpha_2} \left(\frac{p_3}{p_2 + p_3} \right)^{\beta_3}, \end{aligned} \quad (5.2)$$

which gives us the probability of the event if we take $Q = 1$ and gives us the desired expectation E if we take $Q = r_1 + r_2 + \alpha_3 + \beta_3$; here α_3 and β_3 refer to the wins of the last tournament winner (or 3^d contestant) before the first tournament win and between the first and second tournament wins, respectively. For $Q = (r_1 + r_2) + \alpha_3 + \beta_3$ we separate E into 3 parts (T_1 , T_2 and T_3) where T_1 corresponds to $r_1 + r_2$, T_2 to α_3 and T_3 to β_3 . Then the first two T 's are easily seen to be (using the linearity of the expectation operator)

$$\begin{aligned} T_1 = & (r_1 + r_2) CD_{(p_1/p_2, p_3/p_2)}^{(1,1)}(r_1, r_3; r_2); \\ T_2 = & \frac{r_1 p_3}{p_1} CD_{(p_1/p_2, p_3/p_2)}^{(1,1)}(r_1 + 1, r_3 - 1; r_2). \end{aligned} \quad (5.3)$$

The third term T_3 can be written (letting $\gamma_3 = \beta_3 - 1$) as

$$\begin{aligned} T_3 = & \frac{p_3}{p_2} \sum_{\substack{\alpha_2 < r_2 \\ \alpha_3 < r_3}} \left[\begin{matrix} r_1 - 1 + \alpha_2 + \alpha_3 \\ r_1 - 1, \alpha_2, \alpha_3 \end{matrix} \right] (r_2 - \alpha_2) p_1^{r_1} p_2^{\alpha_2} p_3^{\alpha_3} \\ & \times \sum_{\gamma_3 < r_3 - 1 - \alpha_3} \left[\begin{matrix} r_2 - \alpha_2 + \gamma_3 \\ r_2 - \alpha_2, \gamma_3 \end{matrix} \right] \left(\frac{p_2}{p_2 + p_3} \right)^{r_2 + 1 - \alpha_2} \left(\frac{p_3}{p_2 + p_3} \right)^{\gamma_3} \end{aligned} \quad (5.4)$$

and we can now replace the limits of summation (for the two outer sums) by $\alpha_2 < r_2 + 1$ and $\alpha_3 < r_3 - 1$ since the terms added and the terms deleted by this are both zero. Separating T_3 into two terms corresponding to r_2 and $-\alpha_2$, we obtain

$$\begin{aligned} T_{31} &= \frac{r_2 p_3}{p_2} CD_{(p_1/p_2, p_3/p_2)}^{(1,1)}(r_1, r_3 - 1; r_2 + 1); \\ T_{32} &= -\frac{r_1 p_3}{p_2} CD_{(p_1/p_2, p_3/p_2)}^{(1,1)}(r_1 + 1, r_3 - 1; r_2). \end{aligned} \quad (5.5)$$

Since $T_{32} = -T_2$ the final answer is $T_1 + T_{31}$. Hence our result for $\mu_1(2)$ for $k = 3$ and $\gamma = 1$ is the sum of 12 CD -functions and this result is already given in (5.1) except that the result contains an arbitrary k instead of $k = 3$.

For the 3 player example with $r_1 = 2, r_2 = r_3 = 1$ with a general \vec{p} -vector with $p_1 + p_2 + p_3 = 1$ the value of $\mu_1(2)$ is

$$\mu_1(2) = 3 - \frac{p_2 p_3 (2 - p_2 - p_3)}{(1 - p_2)(1 - p_3)}, \quad (5.6)$$

which reduces to $8/3$ for $p_1 = p_2 = p_3 = 1/3$. Since the corresponding value was seen to be $10/3$ for discipline D_0 in (4.8), we see that the savings by using discipline D_1 is $2/3$ of a game on the average.

It should be possible to generalize formula (5.1) so that it holds for any s ($2 \leq s \leq k$), but this has not been done. One minor difficulty is that considerably more notation is required. In both sums the superscripts would be $(s - 1, k - s)$ on the CD -function. The first sum would have $k \binom{k-1}{s-1}$ terms since we sum first on j from 1 to k and then on combinations of $s - 1$ components out of $k - 1$. The second (or inner) summation has in addition to this, a sum over the remaining $k - s$ components of each \vec{a} (or \vec{r}) vector, i.e., it would have $k \binom{k-1}{s-1} (k - s)$ terms. Hence altogether there would be $k \binom{k-1}{s-1} (k - s + 1)$ terms, each with a CD -integral. With the help of TS in (5.2), the derivation would be straightforward but not easy to read.

6. A USEFUL IDENTITY FOR NEGATIVE MULTINOMIAL SAMPLING. In (5.2) we used a triple sum (TS) as a starting point (with $Q = 1$) instead of the double sum (DS) given by

$$DS = \sum_{\alpha_1 \geq r_1} \sum_{\alpha_3 < r_3} \begin{bmatrix} r_2 - 1 + \alpha_1 + \alpha_3 \\ r_2 - 1, \alpha_1, \alpha_3 \end{bmatrix} p_1^{\alpha_1} p_2^{r_2} p_3^{\alpha_3}. \quad (6.1)$$

In this section we show the identity between DS and TS without assuming the discipline D_1 , by first showing that DS is equal to a quadruple sum (QS) and then summing out one of the indices in QS to obtain the desired TS . To prove this identity we note that $\alpha_1 \geq r_1$ indicates that the first player has already reached his quota r_1 when the second player reaches his quota r_2 . Hence we have a common point to break up every sequence into 2 disjoint parts, where the first part goes until the first player reaches r_1 and in that part $\alpha_2 < r_2$ and $\alpha_3 < r_3$. In the second part player 2 gets exactly $r_2 - \alpha_2$ wins (including the last win), player 3 gets $\beta_3 < r_3 - \alpha_3$ and player 1 gets $\beta_1 \geq 0$ wins. Hence we have the result that

$DS = QS$, i.e., that

$$DS = \sum_{\substack{\alpha_2 < r_2 \\ \alpha_3 < r_3}} \left[\begin{matrix} r_1 - 1 + \alpha_2 + \alpha_3 \\ r_1 - 1, \alpha_2, \alpha_3 \end{matrix} \right] p_1^{r_1} p_2^{\alpha_2} p_3^{\alpha_3} \\ \times \sum_{\substack{\beta_1 \geq 0 \\ \beta_3 < r_3 - \alpha_3}} \left[\begin{matrix} r_2 - \alpha_2 - 1 + \beta_1 + \beta_3 \\ r_2 - \alpha_2 - 1, \beta_1, \beta_3 \end{matrix} \right] p_1^{\beta_1} p_2^{r_2 - \alpha_2} p_3^{\beta_3}. \quad (6.2)$$

Using the negative binomial identity $(1 - p)^{-m} = \sum_{\gamma=0}^{\infty} \binom{m + \gamma - 1}{\gamma} p^{\gamma}$ with $m = r_2 - \alpha_2 + \beta_3$, $p = p_1$ and $\gamma = \beta_1$ we sum in (6.2) and easily obtain in (6.2) the desired TS in (5.2). This establishes the identity for those sequences where player 1 reaches r_1 first and player 2 reaches r_2 second; a corresponding result holds for each of the $3(2) = 6$ permutations. This shows that the value of β_1 does not affect any computations for $Q = 1$ in (5.2). The extension to k players is straightforward.

7. THE EXPECTED TOTAL NUMBER OF WINS BY A SPECIFIED PLAYER. If we wait only for the first tournament winner then the expected number of wins by player i , denoted by $E_i(W)$, is

$$E_i(W) = r_i D_{a_i}^{(b)}(\vec{r}_i; r_i) + p_i \sum_{\substack{j=1 \\ j \neq i}}^k \frac{r_j}{p_j} D_{a_j}^{(b)}(\vec{r}_j^{(i)}; r_j + 1) \quad (7.1)$$

where the first term corresponds to player i winning and otherwise (in the summation) player j ($j \neq i$) is the winner. As before $\vec{r}_j^{(i)}$ indicates that r_i is replaced by $r_i - 1$ and if $r_i = 1$ in any term then that term is zero. As before $k = b + 1$. Thus if we have a common $r = 1$ then the first D in (7.1) is the answer and if in addition there is a common p then the answer $1/k = (b + 1)^{-1}$ is clearly correct. The derivation of (7.1) is straightforward and we omit it.

As an example we calculate $E_i(W)$ for $i = 1, 2, 3$ for Huygens problem with $\vec{r} = (2, 3, 4)$ and common $p = 1/3$. For the first term in (7.1), we can also use (3.5), (3.6) and (3.7). For $i = 1, 2, 3$ we obtain

$$E_1(W) = 2(.61865569) + .15500686 + .07133059 = 1.4636488, \quad (7.2)$$

$$E_2(W) = 3(.26748971) + .53497943 + .12620028 = 1.4636488, \quad (7.3)$$

$$E_3(W) = 4(.11385460) + .66666667 + .34156379 = 1.4636488, \quad (7.4)$$

all the same and the sum of these three is indeed the same as in (4.5) namely 4.390947. Since the expected number of wins is $\mu/3$ for each player where $\mu = \mu(1)$ is given by (4.5), they must all be equal. More generally for the i th player the expected number of wins is μp_i , regardless of the r -values ($i = 1, 2, 3$).

The second-moment formula analogous to (7.1) is easily shown to be

$$E_i\{W(W - 1)\} = r_i(r_i - 1) D_{a_i}^{(b)}(\vec{r}_i; r_i) \\ + p_i^2 \sum_{\substack{j=1 \\ j \neq i}}^k \frac{r_j(r_j + 1)}{p_j^2} D_{a_j}^{(b)}(\vec{r}_j^{(ii)}; r_j + 2), \quad (7.5)$$

where $\vec{r}_j^{(ii)}$ is \vec{r}_j with the component r_i replaced by $r_i - 2$. Then if $r_i = 1$ or 2 in any term, that term vanishes. We use (7.5) to get σ_i^2 for $i = 1, 2$ and 3 in Huygen's

example above, where $r_1 = 2$, $r_2 = 3$ and $r_3 = 4$. The three answers are

$$\begin{aligned} E_1\{W(W-1)\} &\approx 1.2373114, E_2\{W(W-1)\} \approx 1.9972565, \\ E_3\{W(W-1)\} &\approx 2.4279836. \end{aligned} \quad (7.6)$$

From these we obtain the three variances $\sigma_i^2 (i = 1, 2, 3)$ as

$$\sigma_1^2 \approx .5586923, \quad \sigma_2^2 \approx 1.3186374, \quad \sigma_3^2 \approx 1.7493645. \quad (7.7)$$

These are increasing with $r_i (i = 1, 2, 3)$ as one might suspect.

In the $k = 3$ example with $r_1 = r_2 = r_3 = 1$, $s = 1$ and common p , the second moments are all zero and hence the three variances are

$$\sigma_i^2 = EW_i - [EW_i]^2 = \frac{1}{3} - \frac{1}{9} = \frac{2}{9} (i = 1, 2, 3). \quad (7.8)$$

8. THE EXPECTED WAITING TIME: SPECIAL CASE OF COMMON ARGUMENTS. The expected waiting time for k players with a common p -value of $1/k$ and with equal or unequal r -values has not been heretofore studied or tabulated, although for equal r -values it is given simply in our notation by a single Dirichlet integral, namely by

$$E(WT) = k^2 r D_1^{(b)}(r; r+1), \quad (8.1)$$

where $b = k - 1$ as before and the scalar arguments (1 and r) are used when the b components of the vectors \vec{a} and \vec{r} , respectively, are all equal. The numerical values (cf. Table below) are remarkably smooth and monotonic both for increasing r and increasing k . The increase with respect to r is obvious (the first differences are actually increasing with r), but the same property also holds with respect to k and this is less obvious (the first differences are actually decreasing with k). A short table for $k = 2, 3, 4, 5$ and 10 is given below for $r = 1(1)10(5)100$.

Although the simple upper bound kr for (8.1) is easy to obtain, it appears to be challenging to get a good asymptotic approximation for r and k both large. If we use kr as a rough approximation for that purpose, the error decreases with r but increases with k . Thus for $r = 100$ the error in using $kr = 200$ for $k = 2$ is 6% but the error in using $kr = 1000$ for $k = 10$ is 15%. We consider a better approximation below for $E\{WT\}$ based on statistical considerations.

Using the result in (2.34b) of [5] with $j = 0$, $a = 1$, $m = r + 1$, namely

$$D_1^{(b)}(r, r+1) = \int_0^\infty G_r^b(x) dG_{r+1}(x), \quad (8.2)$$

where $G_r(x)$ is the incomplete gamma function given by

$$G_r(x) = \frac{1}{\Gamma(r)} \int_0^x t^{r-1} e^{-t} dt, \quad (8.3)$$

it can easily be shown that (8.1) above is $kE(k, r)$ where $E(k, r)$ is the expected minimum of k random variables which are iid, each of which has the gamma distribution with parameter r . Hence the expected minimum of these can be approximated by the $1/(k+1)$ quantile of the common distribution $G_r(x)$, i.e., by the solution in y of

$$G_r(y) = 1 - e^{-y} \sum_{i=0}^{r-1} \frac{y^i}{i!} = \frac{1}{k+1} \quad (8.4)$$

TABLE 8.1. Expected Waiting Time (common r , $p = 1/k$)

r	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 10$
1	1.000000	1.000000	1.000000	1.000000	1.000000
2	2.500000	2.888889	3.218750	3.510400	4.660216
3	4.125000	5.049383	5.864136	6.606313	9.711140
4	5.812500	7.348270	8.730541	10.010009	15.533842
5	7.539063	9.734204	11.736762	13.609403	21.856845
6	9.292969	12.181254	14.841424	17.346933	28.536383
7	11.067383	14.674221	18.020165	21.188563	35.486123
8	12.857910	17.203304	21.257233	25.112238	42.649598
9	14.661530	19.761754	24.541762	29.102717	49.987546
10	16.476059	22.344688	27.865883	33.148950	57.471384
15	25.666067	35.522923	44.912981	53.981108	96.467168
20	34.985173	48.991265	62.429827	75.475620	137.204701
25	44.386241	62.643880	80.245210	97.391287	179.051910
30	53.845310	76.426923	98.271953	119.605204	221.685783
35	63.348217	90.308563	116.458502	142.044658	264.914914
40	72.885770	104.268132	134.771370	164.662406	308.614824
45	82.451599	118.291305	153.187347	187.425703	352.699167
50	92.041076	132.367675	171.689565	210.310745	397.105149
55	101.650713	146.489395	190.265309	233.299584	441.785404
60	111.277802	160.650378	208.904709	256.378292	486.703164
65	120.920194	174.845785	227.599924	279.535801	531.829205
70	130.576146	189.071699	246.344603	302.763146	577.139847
75	140.244223	203.324892	265.133513	326.052940	622.615582
80	149.923227	217.602669	283.962284	349.399015	668.240112
85	159.612146	231.902751	302.827223	372.796156	713.999660
90	169.310113	246.223190	321.725172	396.239906	759.882447
95	179.016382	260.562308	340.653409	419.726422	805.878316
100	188.730304	274.918644	359.609565	443.252360	851.978433

which we denote by $G_r^{-1}(1/k + 1)$. Then our first approximation is

$$kG_r^{-1}\left(\frac{1}{k+1}\right) \approx E\{WT\}. \quad (8.5)$$

For $r = 100$, $k = 2$ this gives 190.859 with error of 1.1% and for $r = 100$, $k = 10$ it gives 869.284 with error of less than 1%; clearly kr gives the upper bounds 200 for $k = 2$ and 1000 for $k = 10$. Inversion of the series in (8.4) would give essentially the same result in the form of an explicit expression, but the result is not simple and the convergence is slow; hence we omit further development along this line. We illustrate the applications of the table 8.1 with the following two problems:

Problem 1. Suppose $k = 3$ players enter into a multinomial tournament consisting of 3-player games, where the first one to win $r = 10$ games wins the tournament. Each player has probability $1/10$ of winning a game and the games are independent of each other; the remaining $7/10$ probability corresponds to draws, i.e., games with no winner. What is the expected number of games needed to obtain a first tournament winner?

We use the fact that 3 players have equal probability and condition on the non-drawn games in which the common conditional $p = 1/3$. Then by using the table for $k = 3$ and $r = 10$ and dividing by $1 - p_d$, where $p_d = .7$ is the probability

of a draw in each game, we obtain

$$E\{WT\} = \frac{1}{0.3}(22.344688) = 74.482293, \quad (8.6)$$

the large answer being due to the high frequency of draws. ■

Problem 2. If $k = 5$ players are assigned to the faces 1, 2, 3, 4 and 5 of a single fair die and the side 6 corresponds to a draw, what is the expected number of tosses needed to obtain a first tournament winner if 100 wins are needed to win the tournament.

Using the same method as in problem 1, we take 443.252360 from the table for $r = 100$ and $k = 5$ and obtain

$$E\{WT\} = \frac{1}{5/6}(443.252360) = 532.903 \quad (8.7)$$

as the expected number of tosses required. ■

If we make use of the fact that $G_r(X)$ as a transformation yields the uniform $U(0, 1)$ distribution, we can obtain a correction term to the approximation in (8.5). Using a standard asymptotic method based on the Taylor series expansion for $G_r^{-1}(x)$, we obtain

$$E\{G_r^{-1}(X)\} \approx G_r^{-1}\left(\frac{1}{k+1}\right) - \frac{1}{2!} \frac{k}{(k+1)^2(k+2)} \frac{g'_r(y)}{[g_r(y)]^3}, \quad (8.8)$$

where

$$g_r(y) = \frac{y^{r-1}e^{-y}}{\Gamma(r)} \quad \text{and} \quad g'_r(y) = \frac{d}{dy}\{g_r(y)\} \quad (8.9)$$

and y is the solution of (8.4).

To illustrate the numerical results for this asymptotic analysis we give the results for $r = 100$ (the last row of table 8.1) and $k = 2, 3, 4, 5$ and 10. The respective approximations are

$$189.4, \quad 276.0, \quad 361.1, \quad 444.9 \quad \text{and} \quad 854.5. \quad (8.10)$$

The maximum percentage error for these five approximations is 0.4%.

9. CONCLUDING REMARKS. The authors have already noted a massive analogy between multinomial problems (sampling from an infinite population or with replacement) and hypergeometric problems (sampling from a finite population without replacement). All aspects of the sharing problem above can be extended to analogous problems with a finite population when sampling without replacement. For purposes of brevity none of these problems are included in this paper.

There are interesting analogies to be pointed out even when both problems are multinomial. Thus the sharing problem has many analogies with the following so-called “Banach match box” problem [5, p. 69]: A smoker starts with r_i matches in the i th box and selects a box at random (i.e., with equal probabilities for each remaining box) every time he (or she) lights up. There are different disciplines depending on what we do with the emptied match boxes. The discard of the empty match box corresponds exactly to the withdrawal of the early winners from the tournament. Moreover each box emptied corresponds to some player reaching his or her quota and hence if we match up the parameters properly the problems

become identical. However some of the questions asked may be quite different. Thus there is little interest in the vector-ranking of the match boxes. On the other hand, some of our formulas above can be directly applied to the Banach match box problem and the development of variance formulas for the waiting time seems more important there than in the sharing problem.

Finally the sharing problem is analogous to the following version of the birthday (or birthmonth) problem: What is the expected number of people that have to be polled to find r people with the same birthday (or birthmonth or born under the same zodiac sign)? In this case the “contestants” are the days of the year (or the month) and the winner of the tournament is the day (or month) which is the first one to come up with r births in it.

REFERENCES

1. C. H. Huygens (1657), *De Ratiociniis in Ludo Aleae*, 16 pages in appendix to Schooten's book entitled “*Exercitationum Mathematicarum libri quinque*”, Leyden, the Netherlands (1657).
2. K. Jordan (1972), *Chapters on the Classical Calculus of Probability* (p. 440). *Disquisitiones Mathematicae Hungaricae* 4, Akadémiai Kiadó, Budapest, Hungary.
3. L. de B. Pacioli (1494), *Summa de Arithmetica, Geometria, Proportioni e Proportionalita* (p. 197) Venice, Italy.
4. Milton Sobel, V. R. R. Uppuluri, and K. Frankowski (1977), *Selected Tables in Mathematical Statistics, Volume 4*, Dirichlet Distribution-Type 1. Published jointly by the IMS and AMS, Providence, Rhode Island.
5. Milton Sobel, V. R. R. Uppuluri, and K. Frankowski (1985), *Selected Tables in Mathematical Statistics, Volume 9*, Dirichlet Integrals of Type 2 and their Application. Published jointly by the IMS and AMS, Providence, Rhode Island.
6. L. A., Todhunter (1865), *History of the Mathematical Theory of Probability*, Cambridge, England, (reprinted by Chelsea Publ., N.Y., N.Y., (1949)).

Dept. of Statistics & Applied Probability
University of California, Santa Barbara
Santa Barbara, CA 93106
sobel@bernoulli.ucsb.edu

Department of Computer Science
University of Minnesota
Minneapolis, MN 55455
kfrankow@cs.umn.edu

...an understanding in depth of the mathematics of any given period is hardly ever to be achieved without knowledge extending far beyond its ostensible subject matter. More often than not, what makes it interesting is precisely the early occurrence of concepts and methods destined to emerge only later into the conscious mind of mathematicians; the historian's task is to disengage them and trace their influence or lack of influence on subsequent developments.

—A. Weil

What is Teaching?*

Paul R. Halmos

Do you remember the first time you ever taught? The first day I taught was September 18, 1935—which was slightly over 58 years ago, or, if you want to be very precise, exactly 21,303 days ago today. Does that, I wonder, make me the person with the longest teaching experience in this room? Being aware of the length of my servitude, the authorities in charge of our meeting today deduced that I know, or, in any event, I ought to know what teaching is, and they instructed me to tell you.

The course I taught in 1935 was called freshman algebra; its purpose was to reveal the secrets of quadratic equations (for which there was a formula) and parentheses (which were abominable entities and had to be eliminated at the drop of a hat). The course met at 8:00 in the morning, five days a week—yes, five days, Monday through Friday, inclusive; my pay was \$45.00 a month. Incidentally, I was living at the time in an old-fashioned, comfortable, large, 5-room apartment, within five minutes walk of the campus; the rent was \$45.00 a month.

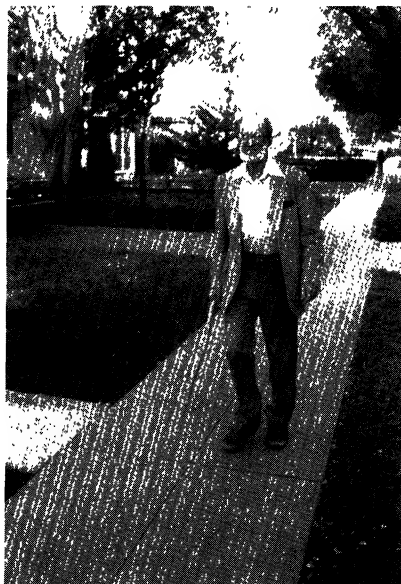
I had no fear of teaching. Stage fright, yes; fear, no. Stage fright, in the sense of being keyed up and slightly nervous is something that has always been with me six minutes for each new class: five minutes before it first meets and one minute after it starts. The same is true of colloquium talks and all other kinds of public appearances.

Even though there was still a lot I had to learn about teaching, I thought I could cope with it. I have always been surprised by beginners who say they don't know how to teach yet. Haven't they already spent almost twenty years under the influence of teachers, and haven't they noticed that some techniques seem to work well and others are just annoying, and haven't they ever muttered to themselves "I could explain that a lot better"? I had had some bad teachers along with the good ones, and I thought I knew what not to do—I marched to my first class confidently with my head held high and with the stage fright barely noticeable under my eagerness to get on with it.

The ages of my students were around 17 and 18; I was a wise old graduate student aged 19. They believed what I told them. Some of them were good and some of them were hopeless. The only one whose name I'll never forget (it was not Drossin, but let's call him that) was one of the hopeless ones. Attendance spotty, homework missing or weak, midterm exams around D minus, the final exam might have helped him pass the course, but I didn't have much faith and neither did he. The Saturday evening before the final a small party at my house was interrupted by the doorbell—it was Drossin. Could he speak to me a minute, privately? Somewhat surprised, I took him to an unused room, and asked him what's up. I was a

*Talk presented at the annual meeting of The Mathematical Association of America, Cincinnati, Ohio, 14 January 1994.

graduate student, wasn't I?, and perhaps I wasn't too well off, was I?, and this course was important for him, so he'd appreciate it if I would help him pass it, and he'd make it worth my while—he'd give me five dollars. Half a week's rent, a week's food! I was too surprised to be angry, but I told him to go away. I returned to the party and told my friends that I had just learned how much I was worth. Next Monday, Drossin's paper was the first I graded. He flunked, with no margin for error.



Paul Halmos engaged in one of his favorite activities.

Well, what about telling you what teaching is? The more I tried to think what to say, the more I was being led to the conclusion that nobody cain't never teach nobody nuttin' nohow. I don't really mean that, but I mean it a lot more than its opposite, and I thought I'd get your attention more quickly by making a crisp statement. I'll spend some of the rest of my time telling you which part of that provocative crispness I really mean.

There are three types of knowledge that we commonly speak of as subjects for teaching or learning; they can be most effectively identified as what, how, and why.

To be educated means to remember something, to be able to use it, and to understand it. Frequently these three kinds of education are thought of as belonging to altogether different kinds of human activity, but ideally they are all present every time. Our memory knows that Napoleon was defeated at Waterloo, our muscles know that certain stretchings and bendings will cause our feet to alternate suitably so as to take us from the office to lunch, and our mind knows why three times five is the same as five times three.

Many students confuse education with memorization. They tend to think that if we know the boiling point of beer, the gestation period of elephants, the conjugation of French irregular verbs, and the population of Burma, together with many other such goodies about the moon, whales, protons, synapses, schizophrenia, and interest rates, then we are educated. A walking encyclopedia is, however, rarely an

educated person. To a historian, history is not just a collection of facts but an organized understanding of how we got to be what we are; Waterloo is not just a fact, but, possibly, a tool to be used to avoid catastrophes in the future. To a chemist, chemistry is not just purple liquids in test tubes, but a scheme for prediction and a way of understanding the world—and the same sort of thing is true of the physicist, the astronomer, the psychologist, and the economist.

Sometimes education is oriented toward application, only application, and the result is just as wide of the mark. A linguist is not one who can speak strange languages, certainly not only that, and a cellist is not one who knows where to put the fingers of his left hand and at what angle to pull his right elbow back. Etymology and syntax for the one, and musicology and musicianship for the other play the role of the vitally necessary components of remembering and understanding.

The third sin, that of identifying education with ratiocination is rarer, but not unknown: the philosophers who, allegedly, don't want to be burdened with or confused by mere facts, and the pure logicians and mathematicians who not only don't want to put their thoughts to work but are even inclined to deny that that can be done, they are the ones guilty of the sin of giving short shrift to matter and muscle.

How then does one learn, and, more to the point here and now, how do we teach the what, the how, and the why? I have my prejudices about all of them, but I can claim professional training and experience about one of them only. I'll wave at the other two, quickly, in passing.

As far as facts go, I am pretty much stumped. How can I learn Napoleon's dates, the meaning of the Hungarian word "mell", the number of rings around Saturn, or the percentage of hydrogen and oxygen in tap water? I can ask an expert, a teacher, by way of a book or a class room, or, when it is physically possible to do so, I myself can look. One trouble is that I might not know enough to look: my teacher has to tell me not only about Napoleon's dates, but also about telescopes. The teaching of "what" abuts on the teaching of "how"—how to learn, how to look, how to perform experiments. The way to teach "what" splits into (1) tell 'em the facts, and (2) tell 'em how to get the facts.

How do we teach the how? How do we teach someone to swim, to play a musical instrument, or to speak a foreign language? One possible answer is: nohow. Don't do nuttin': just wait. Throw the kid into the water, put him on the piano bench, or abandon him in France, and go away. After all, humanity has fumbled its way to these things without any external guidance, and, arguably, the best way for an individual to learn them is to rediscover them for himself. (Incidentally, as far as language goes, is this idea related to Chomsky's innate grammar?)

A somewhat different attitude to how-teaching is to regard the role of the teacher as that of a coach. To be sure, nobody can swim for me, nobody can play the piano with my fingers, and nobody can speak French for me, but somebody can save me an awful lot of time by showing me the right way right away. Once I have seen the crawl, heard the difference that fingering can make to the sound of a piece, or pronounced "an" while holding my nose, I have made hundreds of years of progress.

Does it help the student of swimming, of piano, of French to "understand" what he is doing? Some argue that it does not, that it hurts. (Once you start thinking about how you swim, how fast you should play a passage, or whether you should use the subjunctive, you are lost.) Others argue that all knowledge helps:

the swimmer should understand the pertinent principles of physics, the pianist should see what the theory of harmony has to do with what he is doing, and the speaker should know grammar. You see where the latter view leads, don't you? It says, in effect (and I am inclined to agree), that the teaching of how abuts on the why.

How, finally, do we teach why? How do we teach logic and mathematics, how do we teach abstract concepts and the relations among them, how do we teach intuition, recognition, understanding? How do we teach these things so that when we are done our ex-student can not only pass an examination by naming the concepts and listing the relations, but he can also get pleasure from his insight, and, if he is talented and lucky, be vouchsafed the discovery of a new one? The only possible answer that I can see is: nohow. Don't do nuttin'; just wait. The only way I know of for an individual to share in humanity's slowly acquired understanding is to retrace the steps. Some old ideas were in error, of course, and some might have become irrelevant to the world of today, and therefore no longer fashionable, but on balance every student must repeat all the steps—ontogeny must recapitulate phylogeny every time.

What then can we do to earn a living? Can a mathematician of today, for instance, be of any use to the budding mathematicians of tomorrow? My answer is yes. What we can do is to point a student in the right directions, challenge him with problems, and thus make it possible for him to "remember" the solutions. Once the solutions start being produced, we can comment on them, we can connect them with others, and we can encourage their generalizations. The worst we can do is to give polished lectures crammed full of the latest news from fat and expensive scholarly journals and books—that is, I am convinced, a waste of time.

You recognize, of course, that I seem to be advocating what is sometimes called the Socratic or do-it-yourself or discovery method, or, especially in Texas, the Moore method. The method is not to tell students but to ask them, and, better yet, to inspire them to ask themselves—make students solve problems, and better yet, train students, by example, encouragement, and generous reinforcement, to construct problems of their own. Problem solving—that is the most highly touted current shibboleth, and that is the flag that I too want to wave. The flag should be kept waving; the important ideas deserve to be emphasized over and over again.

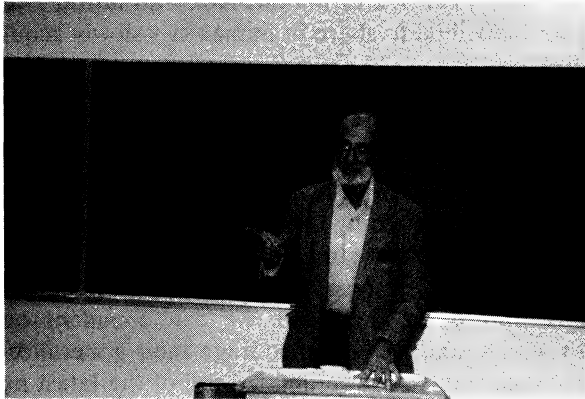
The most effective way to teach mathematics by problem solving is to keep challenging students with problems that are just barely within their reach. One way to inculcate the historical attitude, for instance, is to ask a question that Archimedes didn't have the most efficient tools to answer, and dare students to rediscover Archimedes's research. The best meaning to give to the phrase "undergraduate research" in mathematics is to guide an undergraduate to re-do the research of Leibniz (or Lefschetz).

Everybody loves a puzzle. The ones that appear in the Sunday supplement of the local newspaper or that the telephone company sometimes sends along with their bill, get read and discussed almost as much as the comics and the sports page. The most popular and most widely read part of the *American Mathematical Monthly* is its problem department. Problems is the way to go.

People much prefer stimulation to inundation. Don't snow them; tease them. Puzzles, yes—preachments, no. The problem method of teaching is the best for students, and, once its technical difficulties are overcome, offers the greatest stimulation and reward to the faculty also.

Have students sometimes asked you, when you whizzed through the derivation of the quadratic formula (or the quotient rule for derivatives, or the triangulariza-

tion of complex matrices), possibly in a tone of grudging admiration, “How do you remember all that stuff?” The answer, of course, is that you don’t remember it: you understand it. If students were guided through research that leads them to discover that completing the square does something useful to the equation $2x^2 + 9x + 10 = 0$, they’d have a chance to understand it much better than if they were just shown the technique and then made to practice it a hundred times. The problem method is, I am convinced, the way to teach everything. It teaches technique and understanding, it teaches research and problem solving, it teaches the way nature taught us (about fire and carpentry and the stars and weaving) before we invented teachers.



Paul Halmos.

The method doesn’t begin by proving Theorem 1. It begins with questions: what is true?, what do the examples we can look at suggest? It doesn’t say “look, here’s how it’s done”, it asks “how can it be done?”. It teaches the right attitude toward the solution of all problems. The problems that we came here to solve can be solved—all problems can be solved—and the teaching of problem solving is the way to set about solving them. If we could teach every teacher to teach every course as a problem course, then one generation from now, in twenty five years, say, the need for talks such as this one would no longer exist. All we have to do is find out how to do that, and we can adjourn.

Let me emphasize one thing I casually dropped along the way: I spoke of the way to begin. The way to begin all teaching is with a question. I try to remember that precept every time I begin to teach a course, and I try even to remember it every time I stand up to give a lecture—and, you may recall, I remembered it and acted on it today.

Another part of the idea of the method is to concentrate attention on the definite, the concrete, the specific. Once a student understands, really and truly understands, why 3×5 is the same as 5×3 , then he quickly gets the automatic but nevertheless exciting and obvious conviction that “it goes the same way” for all other numbers. We all seem to have an innate ability to generalize (shades of Chomsky again?). The teacher’s function is to call attention to a concrete special case that hides (and, we hope, ultimately reveals) the germ of the conceptual difficulty.

One time I used the so-called Moore method in an honors class in linear algebra with about 15 students. The first day of class I handed each student a set

of 19 pages stapled together and I told them that they now held the course in their hands. Those 19 pages contained the statements of fifty theorems, and nothing else. There were no definitions, there was no motivation, there were no explanations—nothing but fifty theorems, stated correctly but brutally, with no expository niceties. That, I told them, is the course. If you can understand, state, prove, exemplify, and apply those fifty theorems, you know the course, you know everything that this course is intended to teach you.

I will not, I told them, lecture to you, and I will not prove the theorems for you. I'll tell you, bit by bit as we go along, what the words mean, and I might from time to time indicate what this subject has to do with other parts of mathematics, but most of the classroom work will have to be done by you. I am challenging you to discover the proofs for yourselves, I am putting you on your honor not to look them up in a book or get outside help in any other way, and then I'll ask you to present in class the proofs you have discovered. The rest of you, the ones who are not doing the presenting, are supposed to stay on your toes mercilessly—make sure that the speaker gives a correct and complete proof, and demand from the speaker whatever else is appropriate for understanding (such as examples and counterexamples.)

They stared at me, bewildered and upset—perhaps even hostile. They had never heard of such a thing. They came here to learn something and now they didn't believe they would. They suspected that I was trying to get away with something, that I was trying to get out of the work I was paid to do. I told them about R. L. Moore, and they liked that, that was interesting. Then I gave them the basic definitions they needed to understand the statements of the first two or three theorems, and said "class dismissed".

It worked. At the second meeting of class I said, "O.K., Mr. Jones, let's see you prove Theorem 1", and I had to push and drag them along before they got off the ground. After a couple of weeks they were flying. They liked it, they learned from it, and they entered into the spirit of research—competition, discouragement, glory, and all.

If you are a teacher and a possible convert to the Moore method, don't make the mistake that my students made: don't think that you, the teacher, will do less work that way. It takes me a couple of months of hard work to prepare for a Moore course, to prepare the fifty theorems, or whatever takes their place. I have to chop the material into bite-sized pieces, I have to arrange it so that it becomes accessible, and I must visualize the course as a whole—what can I hope that they will have learned when it's over? As the course goes along, I must keep preparing for each meeting: to stay on top of what goes on in class, I myself must be able to prove everything. In class I must stay on my toes every second. I must not only be the moderator of what can easily turn into an unruly debate, but I must understand what is being presented, and when something fishy goes on I must interrupt with a firm but gentle "Would you explain that please?—I don't understand."

Let me conclude by calling attention to a curious aspect of what I am recommending, an aspect visible in my urging attention to the concrete special case in order to understand the sweeping broad generalization. In effect I am saying that we do not, we cannot understand a vacuum—what we understand is always, in a sense, a fact—and, therefore, just as what cannot be taught without how, and how cannot be taught without why, the question has come full circle around to its start, and it turns out that why cannot be taught without what.

Facts, methods, and insights—all are essential to all of us, all enter all our subjects, and our principal job as teachers is to sort out the what, the how, and the

why, point the student in the right direction, and then, especially when it comes to the why, stay out of his way so that he may proceed full steam ahead.

Having expressed my strong views about why we should teach problem courses, I call your attention to my major omission: I haven't said a single word about how to do that. How, exactly, does one go about teaching a problem course in freshman calculus, or, for that matter, in freshman rhetoric, or junior history, or graduate astronomy? I don't pretend to know all the answers, but I have been working at finding them for many years. If you extended my lecture time by another hour or so, or, to be more realistic, by another month or so, I could try to tell you about some of the techniques that I have luckily blundered into. My presentation today was intended to touch briefly on the "why" of such teaching, not the "how". That will have to wait for our next meeting—just tell me when and where it is, and I'll start packing my bag right away.

*Department of Mathematics
Santa Clara University
Santa Clara, CA 95053-2999
Phalmos@SCUACC.SCU.EDU*

An Exchange of Letters

Dear Ms. vos Savant:

This refers to your piece on the Fermat problem.

Mathematicians do not reject a hyperbolic method of squaring the circle, but accept it. Although squaring the circle is not possible in euclidean geometry, it *is* possible in hyperbolic geometry. Both these statements are true: Truth in mathematics is relative to the axiom system. An (imperfect) analogy: although the French say "la lune" where we say "the moon", both terms are correct. Correctness in diction is relative to the language.

Other weak spots in your piece are the suggestions that Wiles's proof is "hyperbolic" and that elliptic geometry might somehow be "in error." Given that the appreciation of mathematics by our citizenry is already fragile and tottering, there is no need to advertise your own misconceptions. Is it truly amazing that a conclusion about ordinary numbers . . . is derived from arguments in geometry? A hint as to how this can possibly be would help readers see that mathematics is a vibrant subject with an elegant and beautiful internal structure.

Sincerely yours,

Leonard Gillman

Dear Dr. Gillman: -

I've read your letter and quite a few others that essentially agree with your conclusion, albeit through very different chains of reasoning, many of them inherently contradictory. I'm sure my mathematician friends would be surprised to read them. They are an education in themselves, and I feel fortunate to be a witness to what has unexpectedly become a fascinating look into the private world of the modern mathematical minds.

I appreciate them all, I learn from them all, and I thank you all for taking the time to write.

Sincerely,

Marilyn vos Savant

What Is Wrong With the Definition of dy/dx ?

Hugh Thurston

We shall use the notations dy/dx and $f'(x)$ freely and interchangeably. [1]

The fact that dy/dx and $f'(x)$ are not interchangeable is evident when you consider that one does not write $dy/d3$ for $f'(3)$. [2]

For a start, the definition is incomplete. It is always a good idea to know what we are talking about, but definitions of dy/dx do not say what the x and y are; in contrast, definitions of $f'(x)$ make it clear that f is a function and x a number.

Secondly, the definition is ambiguous. Most texts describe dy/dx as another notation for $f'(x)$ where $y = f(x)$. For this to be valid they should prove that if $f(x) = g(x)$ then $f'(x) = g'(x)$, but they don't. Indeed, it is hard to *prove* anything about dy/dx without knowing what x and y are. All we can say for certain is that they are not numbers: $f'(3)$ cannot be denoted by $dy/d3$.

Ambiguity breeds paradox. Whatever x may be, it is something that has values: in the familiar formula $dy/dx|_{x=c}$, c is a value of x . Moreover, y can be constant; we all know that if y is constant then dy/dx is zero. The formula

$$\frac{dx}{dy} = 1 \bigg/ \frac{dy}{dx}$$

shows that x and y are the same kind of entity, so in principle x can be constant. Suppose that x is constant with value 1. Then $x^2 = x^3$, $dx^2/dx = dx^3/dx$, and $2x = 3x^2$. But $2x$ has value 2 and $3x^2$ has value 3. The obvious objection is that it is nonsense to differentiate with respect to a constant—we cannot have a rate of change with respect to something that is not changing. This objection may be obvious, but it is not valid; if x cannot be constant in dy/dx there should be something in the definition that implies this.

So what are x and y ? We can take a hint from the fact that they have values. There is a familiar entity that has values—the function. Or we can make the reasonable suggestion that the x and y in dy/dx are the same as in dx and dy . In the modern (Fréchet) theory of differentials, x and y are functions. We don't apply Fréchet theory to elementary calculus, but if we did the x and y would be the familiar type of function whose values and arguments are real numbers. From now on by “function” I shall mean this type of function.

Now what is dy/dx ? First, look to Leibniz. His dx was an infinitesimal increment in x . An increment in x is $x(c+h) - x(c)$ and the corresponding increment in y is $y(c+h) - y(c)$. Then the value of dy/dx at c is

$$\frac{y(c+h) - y(c)}{x(c+h) - x(c)} \tag{1}$$

where h is infinitesimal. Nowadays instead of having h infinitesimal we have it approach zero. Then (1) becomes $y'(c)/x'(c)$ if x and y are differentiable at c and $x'(c) \neq 0$.

We find the same result if we consider tangents: the slope of a secant of the graph of y against x is (1), and the slope of the tangent is its limit. Rates of change lead to the same result: if x and y represent two quantities that vary with time the average rate of change of the second quantity with respect to the first between times c and $c + h$ is (1) and the instantaneous rate of change is its limit.

All this suggests the following definition.

Definition. If x and y are functions, dy/dx is y'/x' .

From this definition we can prove the familiar rules of differentiation, no longer as mere rules but as properly-stated theorems. It is convenient to use the not-uncommon figures of speech “ f exists at c ” for “ $f(c)$ exists” and “ $f = g$ at c ” for “ $f(c) = g(c)$ ”.

For example

Theorem (chain rule). If x , y and z are functions,

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx}$$

wherever the right-hand side exists.

Proof: If the right-hand side exists at c , then

$$\left(\frac{dz}{dy} \frac{dy}{dx} \right)(c) = \frac{dz}{dy}(c) \frac{dy}{dx}(c) = \frac{z'(c)y'(c)}{y'(c)x'(c)} = \frac{z'(c)}{x'(c)} = \frac{dz}{dx}(c).$$

We can also say clearly and definitely:

Theorem. If x is constant, dy/dx does not exist anywhere.

Proof: x' has the value 0.

Theorem. If x is increasing on an interval I and dy/dx is negative on I then y is decreasing on I .

There are analogous results if x is decreasing or dy/dx is positive or both. The proofs are obvious.

Our definition legitimizes the use of parametric and implicit differentiation. For example,

$$x(t) = 2 \cos t, y(t) = \sin t$$

is a parametrization of the ellipse $x^2 + 4y^2 = 4$. We have

$$\frac{dy}{dx}(t) = \frac{y'(t)}{x'(t)} = -\frac{\cos t}{2 \sin t},$$

giving the slope at $(2 \cos t, \sin t)$. This calculation, and the corresponding implicit one, occur in texts, but in any text which defines dy/dx only where y is a function of x they are necessarily invalid.

Finally how does our definition relate to the traditional one? First, what is $f(x)$ if x is a function? It is the function $t \rightarrow f(x(t))$. For example, the speed of a body moving with positive acceleration is a function of the distance covered: if y denotes the speed and x the distance, then $y = f(x)$ for some f . (If the acceleration a is constant, $f(x) = \sqrt{2ax}$.) At time t , when the distance is $x(t)$, the speed is $f(x(t))$, so that $y(t) = f(x(t))$.

It follows that if $y = f(x)$ then $y' = f'(x)x'$ wherever the right-hand side exists, and so $dy/dx = f'(x)$ wherever the right-hand side exists and x' has a non-zero value.

Adopting the definition suggested here would not alter the well-known and universally-accepted formulas involving the Leibnizian derivative, but it would give them a sound basis.

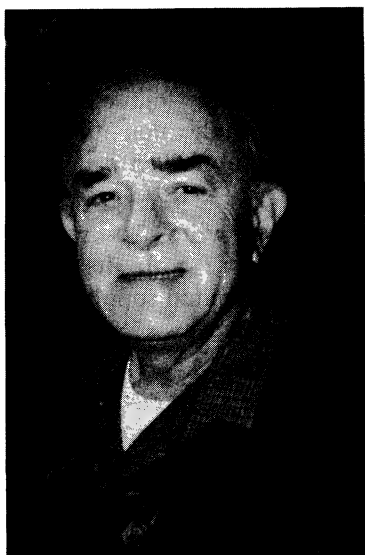
REFERENCES

1. Serge Lang, *A First Course in the Calculus* (third edition) 1971, p. 248.
2. A. R. Pargeter, *Mathematical Gazette*, 54 (1970) p. 165.

*Department of Mathematics
University of British Columbia
#121-1984 Mathematics Road
Vancouver, B.C., Canada V6T 1Y4*

PICTURE PUZZLE

(from the collection of Paul Halmos)



They frequently collaborated, but these photos were taken far apart: the first in 1986 and the second in 1938.

(see page 923.)

the possibility of a Chinese origin for the Hindu-Arabic numeral system (Wylie, 1897). The conjecture was echoed by early 20th century scholars (Fellows, 1921). In most recent times, this claim has been made most succinctly in the works of Wang Ling and Joseph Needham. In a 1958 paper delivered in Adelaide, Wang presented a detailed case for a Sino origin of the “Hindu-Arabic” numerals and pointed to the strong possibility of westward transmission to India. Wang’s theory was further amplified in his collaborative work with Joseph Needham. *Science and Civilization in China*, vol. 3, devotes several pages (pp. 146–150) to this very issue and the phenomena of “stimulus diffusion”. Needham’s work clearly indicates the need for further research clarification as to the status of early Hindu mathematics and the possibility of cultural transmissions. It is exactly this research that must be undertaken to strengthen the claim for a Chinese genesis of our numeral system and, unfortunately, it is exactly this research that is lacking in *Fleeting Footsteps*. What was the status of ancient Indian mathematics during the Warring States period of Chinese history? How were the numerals used in ancient India? Could the Chinese have obtained their mathematical knowledge from India?—after all, Buddhism was an intellectual import from China’s western neighbor. These are some of the issues and questions that must be addressed in positing a claim of a Chinese origin for the “Hindu-Arabic” numeral system and they remain missing footsteps in the path this book has taken.

Despite the inability to develop and strengthen its major premise, *Fleeting Footsteps* is a valuable resource for understanding early Chinese mathematics. Its sections devoted to the *Sun Zi suanjing* and its problems and to rod calculations provide historical information and insights previously unknown to a large audience. Although the “Fleeting Footsteps” themselves remain elusive, the book that bears this name is highly recommended for general reading and library acquisition.

REFERENCES

- Cajori, Florian, *A History of Elementary Mathematics*, New York: Macmillan Co., (1896).
 Eastlake, F. W., Finger Reckoning in China, *China Review* 9 (1880) 250–251.
 Fellows, Albion, Chinese mathematics—the Oldest, *Chinese Review* 1 (1921) 214–215.
 Needham, Joseph, *Science and Civilization in China* vol. 3, Cambridge: Cambridge University Press, (1959).
 Swetz, Frank, The evolution of mathematics in ancient China, *The Mathematics Magazine*, 52 (1987) 10–19.
 Wang Ling, The Chinese origin of the decimal place—value system in the notation of numbers and the possibility of its transmission to India, Paper presented at International Congress, Adelaide, Australia, (20–27 August, 1958) 5.
 Wylie, Alexander, *Chinese Researches*, Shanghai: Mission Press, (1897).

Department of Mathematics
Pennsylvania State University
Middletown, PA 17057-4898

Answer to Picture Puzzle
(p. 857)

Philip Harman and Aurel Wintner

A Stochastic Approach to the Gamma Function

Louis Gordon

1. INTRODUCTION. A measure of the gamma function's importance and ubiquity is that Erdelyi et al. (1981) give it pride of place in their encyclopedic treatment of special functions. Davis (1959) places the gamma function in a masterful historical context. Artin (1964) develops its elementary theory as a function on the real line in a beautiful exposition whose unifying theme is the solution of functional equations. Whittaker and Watson (1927) provide an extensive treatment of the gamma function of complex argument.

Our purpose in this study is to provide a rigorous and unified exposition of the theory of the gamma function by elementary probabilistic means. Because an understanding of the gamma function is crucial to understanding much of statistical theory, such a treatment should be useful. In the process, we also illustrate a useful style of argument—to obtain a result about distributions, realize random variables in the most convenient fashion, evaluate probabilities for these variates, then express the results in terms of distributions, where the variates' provenance is irrelevant. We also hope that some of the bounds we develop in discussing the Stirling formula may be of interest in their own right.

We start by defining the gamma function for $t > 0$ as Euler's second integral: $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$. That $\Gamma(t)$ satisfies the fundamental functional equation $\Gamma(t+1) = t\Gamma(t)$ follows immediately from integration by parts. That the normal probability density $(2\pi)^{-1/2} e^{-x^2/2}$ integrates to 1 allows us to evaluate $\Gamma(1/2) = \sqrt{\pi}$. From this probabilistic foundation will rapidly follow the Legendre duplication formula, Stirling's formula, Gauss' multiplication theorem, the Newman product, the reflection formula and Euler's sine product.

2. THE GAMMA FUNCTION FOR POSITIVE t . We write X_t for a gamma-distributed random variable with shape parameter $t > 0$. Its density is $x^{t-1} e^{-x} / \Gamma(t)$ for $x \geq 0$. We assume we are given an unlimited supply of independent unit exponential random variables Y_j , each having density e^{-x} on the positive half line. Given random variables S and T , we write $S \stackrel{\text{stoch}}{=} T$ for stochastic equality, namely that the two random variables have the same distribution.

We formulate many results in terms of the logarithms of gamma variates. The following technical lemma is useful in dealing with various remainder terms. We then give the identity on which our entire presentation is based.

Lemma 1. *If X_t is gamma-distributed with shape $t > 0$ then $E\{\log(X_t/t)\} \leq 0$. If $t > 1$ then $E\{\log(X_t/t)\} > -1/(t-1)$. In addition,*

$$E\{\log^2(X_t/t)\} < 2t/(t-2)^2 \text{ if } t > 2. \quad (1)$$

Proof: For the first two inequalities, apply $1 - x^{-1} \leq \log x \leq x - 1$, valid for $x > 0$, to $\log(X_t/t)$, and take expectations. For the last, use $\log^2 x < (1 - x^{-1})^2 + (x - 1)^2$. \square

We now present the fundamental representation in Theorem 2. The second identity (3) is too pretty and suggestive to pass up. In reality, when we have to rearrange series or compute expectations, we always go back to (2), where the real dirty work gets done.

Theorem 2. *The logarithm of a gamma-distributed variate X_t with shape parameter t can be decomposed into a sum of independent unit exponential variates Y_j and an independent remainder term. Specifically, for $n \geq 1$,*

$$\log(X_t) \stackrel{\text{stoch}}{=} -\gamma + \left[\sum_{j=0}^{n-1} \frac{1}{j+1} - \frac{Y_j}{j+t} \right] + \log(X_{t+n}/(t+n)) + d_{n,t} \quad (2)$$

$$\log(X_t) \stackrel{\text{stoch}}{=} -\gamma + \sum_{j=0}^{\infty} \frac{1}{j+1} - \frac{Y_j}{j+t} \quad (3)$$

where X_{t+n} is gamma-distributed with shape $t+n$, independent of the other summands in the representation, γ is the Euler constant, and the scalar remainder $|d_{n,t}| < (t+1)/n$.

Proof: We use the elementary fact that if X_t is gamma-distributed and Y is independently exponential then $X_t + Y$ is gamma-distributed with shape $t+1$, independent of the ratio $R = X_t/(X_t + Y)$, which has density $f_R(r) = tr^{t-1}$ concentrated on the interval $(0, 1)$. Directly evaluate the distribution function to show that R^t has a uniform distribution.

Let $S_0 = 0$ and $S_j = \sum_{i=1}^j Y_i$. For any $n \geq 1$,

$$\log(X_t) = \log\left(\prod_{j=1}^n \frac{X_t + S_{j-1}}{X_t + S_j}\right) + \log(X_t + S_n) \stackrel{\text{stoch}}{=} \sum_{j=1}^n \log(U_j^{1/(t+j-1)}) + \log(X_{t+n}),$$

where U_1, U_2, \dots stand for independent uniform variates on $(0, 1)$. Denote by h_n the sum of the first n terms in the harmonic series. Observe that $-\log(U_j) \stackrel{\text{stoch}}{=} Y_j$, so that

$$\begin{aligned} \log(X_t) \stackrel{\text{stoch}}{=} & \log(t+n) - h_n \\ & + \sum_{j=0}^{n-1} \left[(j+1)^{-1} - Y_j(j+t)^{-1} \right] + \log(X_{t+n}/(t+n)). \end{aligned}$$

Let $d_{n,t} = \gamma + \log(t+n) - h_n$, with limit 0 as $n \rightarrow \infty$. Note $|d_{n+1,t} - d_{n,t}| < (t+1)/(n(n+1))$. The bound is a term in a telescoping series, proving (2).

The remainder $\log(X_{t+n}/(t+n))$ is negligible in mean-square because of (1). Assertion (3) follows from the Chebychev inequality. \square

We now use the representation to derive many of the important properties of the gamma function. Start by observing that the first and second logarithmic derivatives of the gamma function—the digamma and trigamma functions—are

respectively the expectation and variance of $\log X_t$:

$$E\{\log X_t\} = \int_0^\infty \log(x) x^{t-1} e^{-x} dx / \Gamma(t) = \frac{d}{dt} \log \Gamma(t) = \psi(t) \quad (4)$$

and

$$\text{Var}\{\log X_t\} = \frac{\int_0^\infty \log^2(x) x^{t-1} e^{-x} dx}{\Gamma(t)} - \left[\frac{\int_0^\infty \log(x) x^{t-1} e^{-x} dx}{\Gamma(t)} \right]^2 = \psi'(t). \quad (5)$$

The following identities are of fundamental importance. The first is due to Wilks (1932), the second is the Legendre duplication formula.

Theorem 3. For $t > 0$, let X_t and $X_{t+1/2}$ be independent gamma variates. Then

$$2^2 X_t X_{t+1/2} \stackrel{\text{stoch}}{=} X_{2t}^2 \quad (6)$$

$$2^{2t-1} \Gamma(t) \Gamma(t+1/2) = \Gamma(2t) \Gamma(1/2). \quad (7)$$

Proof: As in the classical proofs, break (2) into a stochastic piece of odd-numbered terms, a stochastic piece of even-numbered terms, and a scalar remainder. Next take limits to obtain

$$2 \log(X_s) \stackrel{\text{stoch}}{=} \log(X_{s/2}) + \log(X_{(s+1)/2}) + 2 \log(2).$$

Wilks' theorem follows by the substitution $2t = s$. Finally, set $t = 1/2$, raise both sides of (6) to the r power, and take expectations to obtain

$$\frac{\Gamma(2r+1)}{\Gamma(1)} = E\{X_1^{2r}\} = 2^{2r} E\{X_{1/2}^r\} E\{X_1^r\} = 2^{2r} \frac{\Gamma(r+1/2)}{\Gamma(1/2)} \frac{\Gamma(r+1)}{\Gamma(1)}.$$

The assertion follows because $\Gamma(r+1) = r\Gamma(r)$. \square

Next is Stirling's formula. With little effort we provide inequalities instead of the usual asymptotics.

Theorem 4. Let X_t be gamma-distributed with shape parameter t . For all $0 < t < \infty$, $\psi'(t) = \text{Var}\{\log(X_t)\} = \sum_{j=0}^\infty (t+j)^{-2}$ and $\psi(t) - \log(t) \rightarrow 0$ as $t \rightarrow \infty$. In addition,

$$\left\{ \frac{1}{t} + \frac{1}{2t^2} + \frac{1}{6(t+1/4)^3} \right\} < \psi'(t) < \left\{ \frac{1}{t} + \frac{1}{2t^2} + \frac{1}{6t^3} \right. \\ \left. \frac{1}{t} + \frac{1}{2t^2} + \frac{1}{6t^3} - \frac{1}{30t^5} \right\} \quad (8)$$

Proof: Compute the moments of $\log(X_t)$ from (2), (4), (5), Lemma 1 and Theorem 2, noting the latter involves independent variates, each with variance 1 and an independent remainder term with negligible variance. The proof of (8) is motivated by the desire to approximate the variance with an integral by means of the

trapezoid rule. Formally, write

$$\begin{aligned}\psi'(t) &= \frac{1}{2t^2} + \frac{1}{t} + \frac{1}{2} \sum_{j=0}^{\infty} \left\{ \frac{1}{(t+j)^2} + \frac{1}{(t+j+1)^2} - 2 \left[\frac{1}{t+j} - \frac{1}{t+j+1} \right] \right\} \\ &= \frac{1}{t} + \frac{1}{2t^2} + \frac{1}{2} \sum_{j=0}^{\infty} \frac{1}{(t+j)^2(t+j+1)^2}.\end{aligned}\quad (9)$$

Take $\xi > 0$ and verify by subtraction the two inequalities

$$((\xi + 1/14)^{-3} - (\xi + 15/14)^{-3})/3 < \xi^{-2}(\xi + 1)^{-2} < (\xi^{-3} - (\xi + 1)^{-3})/3,$$

the first by showing $10218656\xi^3 + 1240288\xi^2 - 2087596\xi + 453600 > 0$. The top pair of inequalities in (8) follow from (9). The bottom ones follow similarly from

$$\begin{aligned}- (\xi^{-5} - (\xi + 1)^{-5})/15 &< \xi^{-2}(\xi + 1)^{-2} - (\xi^{-3} - (\xi + 1)^{-3})/3 \\ &< -((\xi + 1/8)^{-5} - (\xi + 9/8)^{-5})/15. \quad \square\end{aligned}$$

We now prove bounds which include Stirling's formula, (11), as a special case.

Theorem 5. For all $t > 0$,

$$\left. \begin{aligned} \log(t) - \frac{1}{2t} - \frac{1}{12t^2} \\ \log(t) - \frac{1}{2t} - \frac{1}{12t^2} + \frac{1}{120(t+1/8)^4} \end{aligned} \right\} < \psi(t) < \left. \begin{aligned} \log(t) - \frac{1}{2t} - \frac{1}{12(t+1/14)^2} \\ \log(t) - \frac{1}{2t} - \frac{1}{12t^2} + \frac{1}{120t^4} \end{aligned} \right\}.\quad (10)$$

In addition,

$$\left. \begin{aligned} \frac{1}{12t + 6/7} \\ \frac{1}{12t} - \frac{1}{360t^3} \end{aligned} \right\} < \log\left(\frac{\Gamma(t)}{\sqrt{2\pi} t^{t-1/2} e^{-t}}\right) < \left. \begin{aligned} \frac{1}{12t} \\ \frac{1}{12t} - \frac{1}{360(t+1/8)^3} \end{aligned} \right\}.\quad (11)$$

Proof: From Theorem 4,

$$\int_t^{\infty} \frac{1}{2x^2} + \frac{1}{6(x+1/14)^3} dx < \int_t^{\infty} \psi'(x) - \frac{1}{x} dx < \int_t^{\infty} \frac{1}{2x^2} + \frac{1}{6x^3} dx,$$

proving the upper pair of inequalities in (10). The lower pair are proved similarly.

To prove (11), we integrate again. We may conclude from (10) that for fixed t the integral $\int_t^y \psi(x) - [\log(x) - (2x)^{-1}] dx$ converges monotonically to a finite limit, say c , as $y \rightarrow \infty$. Hence, the upper pair of inequalities in (10) yields

$$c + \frac{1}{12t + 6/7} < \log\left(\frac{\Gamma(t)}{t^{t-1/2} e^{-t}}\right) < c + \frac{1}{12t}.\quad (12)$$

Similarly integrating the lower pair of inequalities in (10) yields the lower pair in (11).

We now use the Legendre duplication formula (7) to show $e^c = \sqrt{2\pi}$. From (12),

$$\begin{aligned} e^c &= \lim_{t \rightarrow \infty} \frac{\Gamma(2t)}{(2t)^{2t-1/2} e^{-2t}} \\ &= \lim_{t \rightarrow \infty} \frac{1}{\sqrt{\pi}} \frac{\Gamma(t)}{t^{t-1/2} e^{-t}} \frac{\Gamma(t+1/2)}{(t+1/2)^t e^{-(t+1/2)}} \frac{2^{2t-1} e^{-1/2}}{2^{2t-1/2} (1+1/(2t))^{-t}} \\ &= \frac{e^{2c}}{\sqrt{2\pi}}. \quad \square \end{aligned}$$

We next generalize Wilks' identity (6) from which Gauss' multiplication formula will follow. Note that Gauss' formula (14) specializes to the Legendre duplication formula (7).

Theorem 6. *Let $p \geq 2$ be an integer and $t > 0$. Let $X_t, x_{t+1/p}, \dots, X_{t+(p-1)/p}$ be independent gamma variates. Then*

$$p^p \prod_{i=0}^{p-1} X_{t+i/p} \stackrel{\text{stoch}}{=} X_{pt}^p \quad (13)$$

$$p^{pt-1/2} \prod_{i=0}^{p-1} \Gamma(t+i/p) = (2\pi)^{(p-1)/2} \Gamma(pt). \quad (14)$$

Proof: Argue by sieving as in the proof of (7) to show there exists a constant c for which

$$p \log(X_{pt}) \stackrel{\text{stoch}}{=} -p\gamma + c + \sum_{i=0}^{p-1} \sum_{j=0}^{\infty} \frac{1}{j+1} - \frac{Y_{pj+i}}{j+i/p} \stackrel{\text{stoch}}{=} c + \sum_{i=0}^{p-1} \log(X_{t+i/p}),$$

with independent summands on the right. We take expectations to show $e^c = p^p$. For $t > 0$,

$$\begin{aligned} e^c \prod_{i=0}^{p-1} (t+i/p) &= e^c \mathbb{E} \left\{ \prod_{i=0}^{p-1} X_{t+i/p} \right\} \\ &= \mathbb{E} \{ X_{pt}^p \} = \Gamma(pt+p)/\Gamma(pt) = \prod_{i=0}^{p-1} (pt+i). \end{aligned}$$

We prove (14) by applying Stirling's formula to (13). Set $t = 1/p$. Take expectations to obtain $\mathbb{E} \{ [p^p \prod_{i=1}^p X_{i/p}]^r \} = \mathbb{E} \{ X_1^{pr} \}$, giving $p^{pr} \prod_{i=1}^p \Gamma(r+i/p) = b \Gamma(pr+1)$ for all $r > 0$ and some constant, say b . Take the limit as $r \rightarrow \infty$ to show $b = (2\pi)^{(p-1)/2} p^{-1/2}$. Finally apply $\Gamma(x+1) = x\Gamma(x)$ to each side of the equation, proving (14). \square

The Newman product, used by Weirstrass as his definition of the gamma function, follows from the fundamental representation as the moment generating function of $\log(Y)$. It is of interest that $-\log(Y)$ has a Type I extreme value, or Gumbel, distribution.

Theorem 7. For $t > -1$, we have

$$\mathbb{E}\{e^{t \log(Y)}\} = \Gamma(1+t) = e^{-\gamma t} \prod_{j=1}^{\infty} \frac{1}{1+t/j} e^{t/j},$$

when Y has the unit exponential distribution.

Proof: Note that $\mathbb{E}\{e^{-tY}\} = (1+t)^{-1}$. Use (2), with remainders given there, to write

$$\begin{aligned} \Gamma(t+1) &= \int_0^{\infty} y^t e^{-y} dy = \mathbb{E}\{e^{t \log(Y)}\} \\ &= e^{-\gamma t} \left[\prod_{j=0}^{n-1} \frac{1}{1+t/(j+1)} e^{t/(j+1)} \right] e^{td_{n,1}} \mathbb{E} \left\{ \left[\frac{X_{1+n}}{1+n} \right]^t \right\}. \end{aligned}$$

We now deal with the stochastic remainder term. From the Stirling formula, $\mathbb{E}[(X_m/m)^t] = \Gamma(m+t)/(\Gamma(m)t!) \rightarrow 1$ as $m \rightarrow \infty$, proving the theorem. \square

The logistic density $e^{-x}(1+e^{-x})^{-2}$ arises, for example, when studying stochastic models of choice. Direct calculation shows that the random variable $L = \log(U/(1-U))$ is logistic if U is uniformly distributed on $[0, 1]$. From the representation $U \stackrel{\text{stoch}}{=} Y_1/(Y_1 + Y_2)$, it follows that $\log(Y_1) - \log(Y_2)$ has a logistic distribution. We are therefore led to study the moment generating function for a logistic variate by means of the Newman product. The result is the reflection formula for the gamma function.

Theorem 8. Let Y_1 and Y_2 be independent unit exponential variates. For $0 < t < 1$,

$$\mathbb{E}\{e^{t(\log(Y_1) - \log(Y_2))}\} = \Gamma(1+t)\Gamma(1-t) = \prod_{j=1}^{\infty} \frac{1}{1-t^2/j^2} = \frac{\pi t}{\sin(\pi t)}.$$

Proof: Independence and our definition imply $\mathbb{E}\{e^{t(\log(Y_1) - \log(Y_2))}\} = \Gamma(1+t)\Gamma(1-t)$. The representation as an infinite product is immediate from the Newman product, Theorem 7. The last equality is the Euler sine product, the next theorem. \square

We conclude the proof of the reflection formula by establishing Euler's sine product.

Theorem 9. For all real t , $\prod_{j=1}^{\infty} [1 - t^2/(\pi^2 j^2)] = \sin(t)/t$.

Proof: First observe that to pick a number at random from $[0, 1]$, we might as well pick the first n binary random digits uniformly at random, and then finish the job by picking a number uniformly at random in the interval of length 2^{-n} to the right of the chosen dyadic rational. Formally, let B_j be independent Bernoulli variates taking the values 0 or 1 with probability $1/2$ each. For U uniform on $[0, 1]$, we

represent $U = \sum_{j=1}^n 2^{-j} B_j + 2^{-n} U$. Now symmetrize about 0 to obtain $2U - 1 \stackrel{\text{stoch}}{=} \sum_{j=1}^n 2^{-j} (2B_j - 1) + 2^{-n} (2U - 1)$. The symmetrization makes it easy to compute the characteristic functions of left and right sides, yielding

$$\frac{\sin(t)}{t} = \frac{\sin(2^{-n}t)}{2^{-n}t} \prod_{j=1}^n \cos(2^{-j}t), \quad (15)$$

a truncated form of Vieta's formula. Of course one could just as soon pull (15) out of the air, beginning with $\sin(x)/x$ and iterating with $\sin(x) = 2 \sin(x/2) \cos(x/2)$.

Now rearrange (15) and apply the double-angle formula for the cosine to obtain

$$\frac{\sin(t)}{2^{-n} \sin(2^{-n}t)} \frac{1}{\cos(2^{-n}t)} = \prod_{k=1}^{n-1} \cos(2^k 2^{-n}t) = \prod_{k=1}^{n-1} Q_k(\sin^2(2^{-n}t)), \quad (16)$$

where the functions $Q_k(x)$ are polynomials in x of degree 2^{k-1} . Hence the product is a polynomial in $\sin^2(2^{-n}t)$ of degree $2^{n-1} - 1$. The first $2^{n-1} - 1$ positive zeros of the left-hand side occur at $t = k\pi$ for $k = 1, 2, \dots, 2^{n-1} - 1$. Because the left-hand side of (16) is 1 when $t = 0$, and because $\sin(\cdot)$ is strictly increasing in $(0, \pi/2)$, we see that

$$\frac{\sin(t)}{2^{-n} \sin(2^{-n}t)} \frac{1}{\cos(2^{-n}t)} = \prod_{k=1}^{2^{n-1}-1} \left(1 - \frac{\sin^2(2^{-n}t)}{\sin^2(k\pi 2^{-n})} \right).$$

The latter product converges to $\prod_{j=1}^{\infty} (1 - t^2/(j\pi)^2)$ because $x > \sin(x) > x - x^3/6 > 0$ for $x \in (0, \pi/2)$ implies $\sin^2(2^{-n}t)/\sin^2(k\pi 2^{-n})$ is bounded above by t^2/k^2 when $k < 2^{n-1}$. \square

3. BIBLIOGRAPHIC NOTES. It is curious that the development presented here tends to flow in a backwards direction from customary presentations. For example, Wilks (1932) uses the Legendre duplication formula (7) to prove (6) by calculating moments and then appealing to deep results that tell when moments determine uniquely their distribution. Theorems 2, 3 and 6, with similar proofs, appear in Gordon (1989). Bondesson (1978) obtains (3) using characteristic functions. Our proof is closely related to Olshen and Savage's (1968) theory of generalized unimodality.

The first lower bound of (11) slightly improves Robbins' (1955) useful and pretty lower bound $1/(12t + 1)$. Robbins' proof is for integers only. Diaconis and Freedman (1986) remark that the bound is difficult to generalize to non-integer values. We believe the first lower bound and second upper bound of (11) are new. The second lower bound and first upper bound of (8) are implicit in Fichtenholz (1964), Section 540. The corresponding inequalities in (10) and (11) are there given explicitly in Section 541. The use of the Legendre duplication formula to compute the constant in Stirling's formula appears in Henrici (1991), Section 8.5.

Related is work of Blyth and Pathak (1986) whose proof of the Stirling formula uses gamma variates and the central limit theorem. Other approaches to the Stirling formula are those of Namias (1986) by recursion based on the Legendre duplication formula, and of Berndt (1986) who relates $\log(\Gamma(x))$ to the Hurwitz zeta function; he also proves the Gauss multiplication and reflection formulas.

Elementary proofs of Euler's sine product are in Artin (1964) and Eberlein (1977). Artin analyzes the periodic function $\Gamma(x)\Gamma(1-x)\sin(\pi x)$, after extending the domain of $\Gamma(\cdot)$. Feller (1967) uses Artin's proof to find the constant in

Stirling's formula. Eberlein's (1977) proof, whose roots he traces to Euler, uses a clever approximation of $\sin(x)/x$ by a polynomial whose zeros can be explicitly determined. Our proof is a variant of Eberlein's, using a polynomial whose study is naturally motivated in a probabilistic context. Vieta's formula is the starting point for Kac's (1959) monograph.

REFERENCES

- E. Artin, *The Gamma Function*. Holt, Rinehart and Winston. New York (1964).
 B. C. Berndt, The gamma function and the Hurwitz zeta-function. *Amer. Math. Monthly* 92 (1985) 126–130.
 C. Blyth, and P. K. Pathak, A note on easy proofs of Stirling's theorem. *Amer. Math. Monthly* 93 (1986) 376–379.
 L. Bondesson, On infinite divisibility of powers of a gamma variable. *Scand. Actuarial J.* (1978) pp. 48–61.
 P. J. Davis, Leonhard Euler's integral: A historical profile of the gamma function. *Amer. Math. Monthly* 66 (1959) 849–869.
 P. Diaconis, D. Freedman, An elementary proof of Stirling's formula. *Amer. Math. Monthly* 93 (1986) 123–125.
 W. F. Eberlein, On Euler's infinite product for the sine. *J. Math. Anal. Appl.* 58 (1977) 147–151.
 A. Erdelyi et. al. *Higher Transcendental Functions*, Vol. I. Krieger Publishing Company. Malabar, Fla. (1981).
 W. Feller, A direct proof of Stirling's formula. *Amer. Math. Monthly*. 74 (1967) 1223–1225. Correction *Amer. Math. Monthly* 75 (1968) 518.
 G. M. Fichtenholz, *Differential-und Integralrechnung II*. VEB Deutscher Verlag der Wissenschaften. Berlin (1964).
 L. Gordon, Bounds for the Distribution of the Generalized Variance. *Ann. Statist.* 17 (1989) 1684–1692.
 P. Henrici, *Applied and Computational Complex Analysis*, Vol. 2. John Wiley. New York (1991).
 M. Kac, *Statistical Independence in Probability, Analysis and Number Theory*. Math. Assoc. America. Wash., D. C. (1959).
 V. Namias, A simple derivation of Stirling's asymptotic series. *Amer. Math. Monthly* 93 (1986) 25–29.
 R. A. Olshen, and L. J. Savage, A generalized unimodality. *J. Appl. Prob.* 7 (1968) 21–34.
 H. Robbins, A remark on Stirling's formula. *Amer. Math. Monthly* 62 (1955) 26–29.
 E. T. Whittaker, and G. N. Watson, *A Course of Modern Analysis*, 4th edition. Cambridge University Press. New York (1927).
 S. S. Wilks, Certain generalizations in the analysis of variance. *Biometrika*. 24 (1932) 471–494.

Mathematics Department, DRB-155
University of Southern California
Los Angeles, CA 90089-1113
gordon@mth.usc.edu

An Ode To Fréchet

Some time ago, in Riemann's day,
 Calculus was penned in a cumbersome way;
 seas of partials, indices too,
 Jacobian determinants spiced the stew.
 But now we have another way;
 linear maps have joined the fray.
 "Though some feel wrath, me thinks that Math
 hath truly felt a "breath of Fréchet."

John A. Baker
Pure Mathematics Department
University of Waterloo
Waterloo, Ontario, CANADA
N2L 3G1

Arrangements and Topological Planes

Jacob E. Goodman, Richard Pollack, Rephael Wenger,
Tudor Zamfirescu

1. INTRODUCTION. Let Γ be a finite family of simple curves in the plane. When is there a homeomorphism of the plane to itself that takes all the curves in Γ to straight lines?

In the Euclidean plane, E^2 , we are faced with the fact that two non-intersecting curves in our family must map to two parallel lines. This introduces extraneous technical complications that only distract from the essence of the problem. As with many other geometric questions, it is much simpler to avoid the special cases caused by parallel lines by moving to the projective plane. The real projective plane P^2 is the Euclidean plane E^2 with an extra “line at infinity” adjoined, each point of which represents a parallel direction in E^2 . P^2 has the virtue of simplicity: every pair of points determines a unique line which is topologically a circle (i.e., a simple closed curve), and every two lines meet at a unique point. Thus our question becomes: When is there a homeomorphism of P^2 to itself that simultaneously straightens all the members of a finite family Γ of simple closed curves?

Certainly a necessary condition is that each of the curves be “nicely” embedded in the plane. More precisely, for each curve there must be some homeomorphism that takes P^2 to itself and maps the curve to a straight line. In addition, every two of our curves must meet exactly once, and cross at their point of intersection, just as straight lines do. Are these two conditions sufficient? The answer is no, and a counterexample can easily be constructed using Desargues’ theorem.

Desargues’ theorem, one of the basic theorems of projective geometry, asserts that if the corresponding sides of two triangles meet at three collinear points, then the three lines joining corresponding vertices are concurrent: see Figure 1a. On the other hand, Figure 1b is an example of an arrangement for which Desargues’ theorem fails: any homeomorphism of the plane to itself that mapped the ten curves in Figure 1b to straight lines would yield an arrangement of lines that violated Desargues’ theorem. Hence there is no homeomorphism of the plane to itself that simultaneously straightens the ten curves of Figure 1b.

Let us look for a moment at our two necessary conditions. A straight line l in P^2 does not separate P^2 , since any two points in $P^2 \setminus \{l\}$ are connected by some path, perhaps one crossing the line at infinity. (In contrast to this, a “small” circle does separate P^2 .) Thus if there is to be a homeomorphism of P^2 to itself which maps some simple closed curve l' to a straight line, then l' must also not separate P^2 , i.e., $P^2 \setminus \{l'\}$ must be connected. It follows from Schoenflies’ Theorem (see [12], for example) that the converse is true as well: If l' is a simple curve that does not separate P^2 , then there is a homeomorphism taking P^2 to itself that maps l' to a

Some of the main results of this paper were presented at the Eighth Annual ACM Symposium on Computational Geometry in Berlin on June 11, 1992 [5].

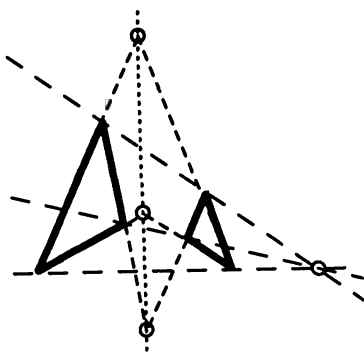


Figure 1a. Desargues' Theorem.

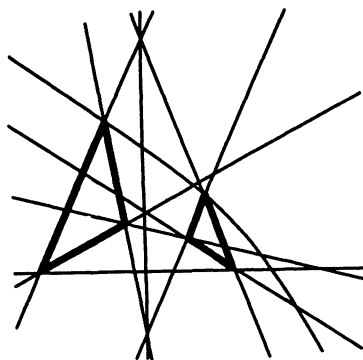


Figure 1b. A non-stretchable arrangement.

straight line. Such a simple closed curve is called a *pseudoline*. A finite family of pseudolines in \mathbf{P}^2 , with the property that any two meet exactly once (and necessarily cross), is known as an *arrangement of pseudolines*.

To visualize an arrangement \mathcal{A} of pseudolines in \mathbf{P}^2 , one can model the projective plane as a circular disk with opposite points identified. For this purpose, remove some pseudoline $l^* \in \mathcal{A}$ from the projective plane. The remaining points then form a space (the Euclidean plane!) homeomorphic to an open circular disk in \mathbf{E}^2 . Call the closure of the disk Δ . Each point on l^* corresponds to an antipodal pair of points on the circle $\partial\Delta$. A pseudoline in \mathcal{A} other than l^* becomes a curve connecting antipodal points in Δ . Thus an arrangement of pseudolines in \mathbf{P}^2 corresponds to a family of Jordan arcs connecting antipodal points on a circle, every pair of arcs intersecting exactly once (or possibly at their endpoints); see Figure 2.

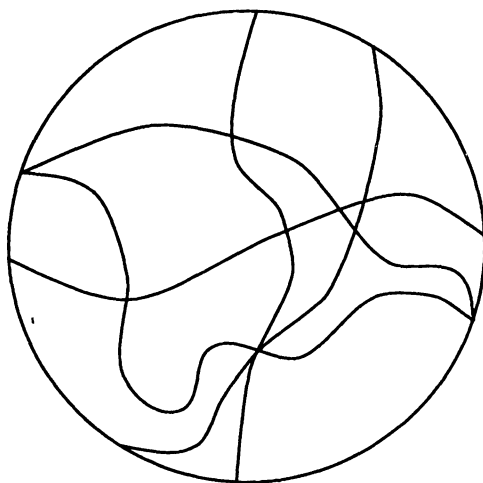


Figure 2. An arrangement of pseudolines in the disk model of \mathbf{P}^2 .

Arrangements of pseudolines have been studied since the work of F. Levi [8], who first pointed out that, in spite of their resemblance to arrangements of straight lines, they are topologically more general objects. B. Grünbaum [6] published an

extensive monograph on arrangements in 1971, in which he answered many questions about line- and pseudoline-arrangements and posed many others. It turns out that the study of arrangements of pseudolines is equivalent to that of oriented matroids of rank 3; see [1] for a definition and good introduction to oriented matroids, and in particular for a discussion of their relationship to pseudoline arrangements.

As in any field of mathematics, we classify arrangements, i.e., we partition them into classes of similar or “isomorphic” ones. This is done by considering their topology: Two arrangements are called *isomorphic* if there is some homeomorphism of \mathbf{P}^2 to itself that maps the pseudolines in one arrangement to those in the other.

An arrangement \mathcal{A} of pseudolines, like an arrangement of straight lines, induces a decomposition of the projective plane into a cell complex $\mathcal{C}(\mathcal{A})$, consisting of cells of dimension 2 (“faces”), 1 (“edges”), and 0 (“vertices”). An isomorphism between two arrangements induces a one-to-one correspondence between their cell complexes that preserves incidence, i.e., neighboring faces in one arrangement map to neighboring faces in the other. The converse is also true: Suppose \mathcal{A} and \mathcal{A}' are arrangements and there is a one-to-one correspondence between $\mathcal{C}(\mathcal{A})$ and $\mathcal{C}(\mathcal{A}')$ that preserves incidence. Patching together homeomorphisms between corresponding faces gives a homeomorphism of the plane to itself that maps the pseudolines of one arrangement to the pseudolines of the other. Thus isomorphism between arrangements is really just a combinatorial relationship that can be defined solely in terms of the cells and their incidences.

The question we posed at the beginning can now be restated as follows: Is every arrangement of pseudolines isomorphic to some arrangement of straight lines? The example above based on Desargues’ theorem shows that the answer is “no”. (G. Ringel showed [15] that even if we consider only arrangements with no multiple points, the answer still remains “no”.) An arrangement of pseudolines that is isomorphic to some arrangement of straight lines is called *stretchable*. Unfortunately, determining whether an arrangement of pseudolines is stretchable turns out to be quite a hard problem (in fact, NP-hard: see [10, 11, 17].)

While not every arrangement of pseudolines is stretchable, arrangements of pseudolines nevertheless share many properties with arrangements of straight lines. Given any arrangement of n straight lines we can always add another line through any two given points not both on the same line to form an arrangement of $n + 1$ straight lines. The same property holds for arrangements of pseudolines [8].

This result, known as the Levi Enlargement Lemma, implies that any arrangement of n pseudolines can be extended to an arrangement of $n + 1$ pseudolines. However, an arrangement of n straight lines can be extended to a much larger and richer structure, the topological space consisting of *all* the lines in the plane. Each line is a “point” in this space, and a neighborhood of the line through two points consists of all the lines through nearby pairs of points.

Can every finite arrangement of pseudolines be embedded in some continuous family of pseudolines analogous to the set of all lines in the plane? Grünbaum posed this question in [6]. What should be the properties of such a family of pseudolines? First, this family should form a topological space analogous to the space of all lines in the plane. Second, this set of pseudolines, together with the set of points in the projective plane, should obey the basic incidence axioms of geometry: any two points should lie on a unique pseudoline, and any two pseudolines should intersect in a unique point. Third, the geometric and topological properties should be linked as they are for lines: as two pseudolines vary continu-

ously, their point of intersection should vary continuously; as two points vary continuously, the unique pseudoline they define should vary continuously as well.

Such structures are known to exist, and have in fact been studied for nearly a century. They are known as *topological projective planes*. A topological projective plane Q , in the sense we will use the phrase, consists of \mathbf{P}^2 as its underlying point set, and a second topological space $L(Q)$ consisting of simple closed curves in \mathbf{P}^2 as its set of “lines”; these satisfy the following conditions:

1. for every two distinct points $p, q \in Q$ there is a unique curve $l(p, q) \in L(Q)$ containing p and q ;
2. every two distinct curves $l, l' \in L(Q)$ intersect in exactly one point at which they cross;
3. $l(p, q)$ varies continuously as a function of p and q ;
4. $l \cap l'$ varies continuously as a function of l and l' .

As before, we use the term *pseudolines* for the curves in $L(Q)$.

Just as the neighborhood of a straight line through two points consists of all the lines through nearby pairs of points, the neighborhood of a pseudoline consists of all pseudolines through nearby pairs of points. Since \mathbf{P}^2 is compact, this is equivalent to the topology induced by the Hausdorff metric on pseudolines: two pseudolines are within distance ϵ of each other if each point of each pseudoline is within ϵ of the other pseudoline; here the metric on \mathbf{P}^2 is the one coming from the standard metric on the sphere S^2 when antipodal points are identified.

Topological planes were discussed by Hilbert in his seminal book *Foundations of Geometry* [7]. There, he gave the first example of a topological Euclidean plane in which Desargues’ theorem failed to hold. F. R. Moulton subsequently gave a simpler example, now known as the Moulton plane [13], which was incorporated into later versions of Hilbert’s book.

More recently, H. R. Salzmann [16] studied topological planes and their various axiomatizations. Among other results, he proved that, with hypotheses even weaker than the above conditions, $L(Q)$ will always be homeomorphic to the space of points of the projective plane \mathbf{P}^2 . He also showed that the fourth condition, that $l \cap l'$ vary continuously as a function of l and l' , is a consequence of the third condition, that $l(p, q)$ vary continuously as a function of p and q , and vice versa. Many other interconnections among properties of topological planes are given in Salzmann’s paper.

Grünbaum’s question can now be reformulated as follows: Given an arrangement \mathcal{A} consisting of a finite number of pseudolines in the projective plane, is there some topological projective plane Q containing \mathcal{A} , i.e., a plane such that $\mathcal{A} \subset L(Q)$? If \mathcal{A} were stretchable, this would be trivially true: take an isomorphic arrangement of straight lines, consider it as embedded in \mathbf{P}^2 , and use the homeomorphism of \mathbf{P}^2 that straightens the pseudolines of \mathcal{A} to define a new topological plane structure on \mathbf{P}^2 in which the “lines” are simply the inverse images of the straight lines of \mathbf{P}^2 . But in general, if \mathcal{A} is not stretchable, no such argument is available. Nevertheless, we will answer the question affirmatively in Section 2 below, by showing how to extend any arrangement of pseudolines to a topological projective plane; the solution will turn out, in fact, to be surprisingly simple.

Grünbaum asked yet another, more sweeping, question in “Arrangements and Spreads”. Assuming that every finite arrangement of pseudolines can be embedded in some topological projective plane, is there a *single* topological plane that contains every finite arrangement of pseudolines up to isomorphism? Another way

of posing this question is to extend our notion of stretchability. Recall that an arrangement of pseudolines is *stretchable* if it is isomorphic to some arrangement of lines in \mathbf{P}^2 . For a topological projective plane Q , let us call an arrangement \mathcal{A} of pseudolines in the projective plane *stretchable in Q* if it is isomorphic to some arrangement of pseudolines $\mathcal{A}' \subset L(Q)$. Then Grünbaum's question becomes: Is there some topological plane Q such that every arrangement of pseudolines is stretchable in Q ?

We will show in Section 3, using the embedding theorem of Section 2, that the answer is again “yes”. Grünbaum called such a topological plane, whose existence he conjectured, a *universal* topological plane.

2. FROM ARRANGEMENTS TO TOPOLOGICAL PLANES. One obvious approach to extending arrangements to topological planes is the repeated use of the Levi Enlargement Lemma, adding new pseudolines one at a time in an infinite process. This would only generate a countably infinite family of pseudolines, however, so one must then “complete” this set by taking some sort of limit. The problem with such an approach is that one may unwittingly introduce discontinuities in taking this limit. While it is conceivable that such a technique may work, no one, to our knowledge, has been able to give a construction along these lines.

Instead of adding pseudolines one at a time, our method will be to define pseudolines piecewise in different regions of the plane, and then to link the pieces together. As previously noted, an arrangement \mathcal{A} of pseudolines splits the projective plane into faces. We will construct the topological plane by defining all the “pseudoline segments” traversing a given face and then showing how to join these segments to form pseudolines with the desired properties.

To carry out our construction, we need a simple fact about stretchability, proved first by J. Richter-Gebert [14] in the “uniform” case, i.e., where no three pseudolines are concurrent:

Lemma 1. *Let \mathcal{A} be an arrangement of n pseudolines. If some face of \mathcal{A} is bounded by at least $n - 1$ pseudolines, then \mathcal{A} is stretchable, i.e., isomorphic to an arrangement of straight lines.*

The lemma is proved, without any assumption of uniformity, at the end of this section.

We now proceed with our construction. Let \mathcal{A} be any arrangement of n pseudolines in \mathbf{P}^2 . Fix some distinguished pseudoline $l^* \in \mathcal{A}$. This pseudoline will play a role in the topological projective plane similar to the role of the line at infinity in the standard model of \mathbf{P}^2 .

For each face f of the arrangement \mathcal{A} , let L_f be the set of pseudolines bounding f . By Lemma 1, $L_f \cup \{l^*\}$ is stretchable. Let h_f be a homeomorphism of the projective plane to itself that maps the pseudolines in $L_f \cup \{l^*\}$ to straight lines (notice that it is possible that $L_f \cup \{l^*\} = L_f$). For each pair of distinct points p and q lying on different segments of ∂f , there is a straight line segment a in $h_f(f)$ connecting $h_f(p)$ to $h_f(q)$. $h_f^{-1}(a)$ is then an arc in f with endpoints p and q ; see Figure 3. (As previously described, the arrangement \mathcal{A} can be visualized as a set of Jordan arcs connecting antipodal pairs of points in a disk where l^* maps to the circle bounding the disk. Of course the “straight” lines shown in Figure 3b are not really *straight* in the disk model; they can, however, be taken to be arcs of circles joining antipodal pairs on the disk boundary—see [6].)

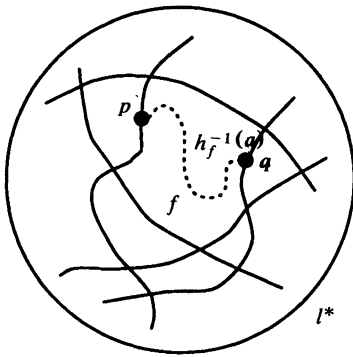


Figure 3a. Original face f .

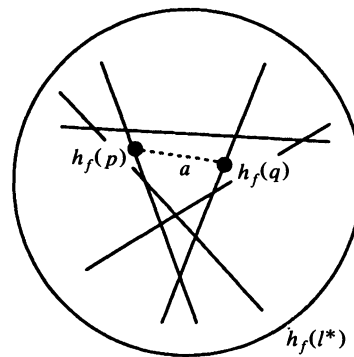


Figure 3b. A straightening of f .

Let Γ be the set of all such arcs over all faces of \mathcal{A} ; the members of Γ form the “pieces” of the pseudolines we are going to construct.

Let γ be some arc of Γ lying in face f . $h_f(\gamma)$ is a line segment which is part of some straight line l meeting $h_f(l^*)$ at a point r “at infinity”. Let $s \in l^*$ be $h_f^{-1}(r)$, and label γ with the point s . (One can think of s as the “slope” of arc γ even though γ may be far from linear, and even though this “slope” depends on the chosen straightening homeomorphism h_f .) Label every arc $\gamma \in \Gamma$ in this manner, and notice that if p is a point on l^* , then every arc with endpoint p has label p . If p is a point on the interior of some edge of f , where $p \in l$ and $l \in L_f$ and $p \notin l^*$, then every point of $l^* \setminus l$ occurs exactly once as a label of an arc in f with endpoint p . Finally, if p is a vertex of f , where $p \in l \cap l'$ and $l, l' \in L_f$ and $p \notin l^*$, then the labels of arcs in f with endpoint p form an arc on l^* between $l \cap l^*$ and $l' \cap l^*$.

The arcs in Γ can now be linked together to form pseudolines. For every point p lying on some pseudoline $l \in \mathcal{A} \setminus \{l^*\}$, let Γ_p be the set of arcs in Γ with endpoint p . Then for every arc $\gamma \in \Gamma_p$ with label s there is exactly one other arc $\gamma' \in \Gamma_p$, on the other side of l , with the same label. Join these two arcs to form a longer arc, and continue. Repeating this for every point p lying on some pseudoline other than l^* , we get a set $\tilde{\Gamma}$ of arcs. We claim that $L(Q) = \tilde{\Gamma} \cup \mathcal{A}$ is the set of pseudolines of a topological plane Q .

For the proof, it is useful to refer again to the disk model described above, in which each point $s \in l^*$ is replaced by an antipodal pair s^+, s^- on the circle $\partial\Delta$ bounding a disk Δ , and a pseudoline that intersects l^* at s becomes a curve with endpoints s^+ and s^- .

We first show that in this disk model the endpoints of every $l \in \tilde{\Gamma}$ constitute an antipodal pair in $\partial\Delta$. Start with any arc $\gamma_0 \in \tilde{\Gamma}$ that forms a segment of l . γ_0 has some label s . Arc γ_0 is separated from s^+ by some $k < n$ pseudolines of \mathcal{A} . One of the two arcs joined to γ_0 is therefore separated from s^+ by only $k - 1$ pseudolines; let this arc be γ_1 . γ_1 also has label s . Repeating this argument k times gives an arc γ_k with label s which has endpoint s^+ ; thus l has s^+ as one endpoint. A similar argument shows that s^- is the other endpoint of l .

If two arcs have the same endpoints on $\partial\Delta$, i.e., the same labels, then it is clear that they can never intersect in the interior of Δ . On the other hand, it is also clear that two arcs in $L(Q)$ cannot intersect infinitely often, since they can meet at most once inside each face of \mathcal{A} . To prove that they intersect *exactly* once, we first establish a general lemma about arcs intersecting in the disk (cf. [4]).

Let l and l' be any two arcs in Δ connecting distinct antipodal endpoints s_1, s_3 and s_2, s_4 , respectively; l and l' may intersect at more than one point. Let p be an isolated point of intersection of l and l' at which they cross. In other words, there is some small topological disk Δ^* containing p and no other point of intersection of l and l' . Arcs l and l' intersect $\partial\Delta^*$ at four points, s_1^*, s_2^*, s_3^* , and s_4^* , with s_i^* lying between p and s_i on l or l' . We say that p is a *proper* intersection point of l and l' if s_1, s_2, s_3 , and s_4 occur in the same order around Δ (clockwise or counterclockwise) as s_1^*, s_2^*, s_3^* , and s_4^* do around Δ^* . (See Figure 4.)

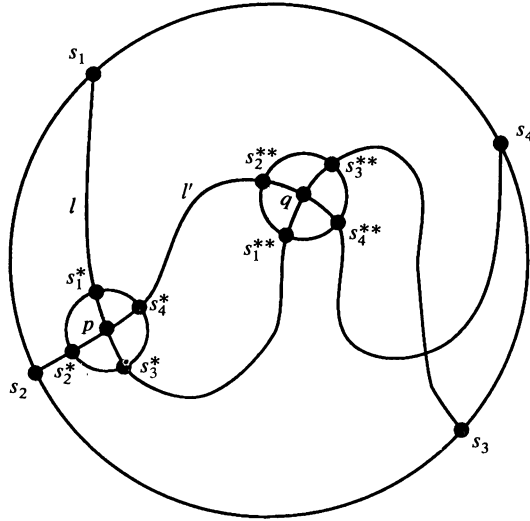


Figure 4. p is a proper intersection point; q is not.

The following lemma replaces the global condition that arcs intersect at precisely one point, at which they cross, by the local condition that every point of intersection is proper.

Lemma 2. *Two arcs connecting distinct antipodal points in the disk that intersect finitely often intersect at precisely one point, and cross there, if and only if every point of intersection of the two arcs is proper.*

Proof: Since the endpoints of each arc are antipodal, the first arc separates the endpoints of the second and thus the arcs must have at least one intersection point. If there is only one, that intersection must clearly be proper. On the other hand, if our two pseudolines l and l' intersect at more than one point, we can list the points of intersection in order along l . Let p and q be two successive points of intersection. Then it follows immediately from the definition that if p is proper, q is not. \square

We now prove that every two arcs in $L(Q)$ intersect exactly once. Let $l, l' \in L(Q)$ be two arcs connecting distinct antipodal endpoints s_1, s_3 and s_2, s_4 ,

respectively. Suppose p is an intersection point of l and l' lying in the interior of a face f . By construction, l and l' meet the boundary of f in four points, s_1^* , s_2^* , s_3^* , and s_4^* , where s_i^* lies between s_i and p on l or l' . The order of s_1 , s_2 , s_3 , and s_4 around Δ agrees with the order of $h_f(s_1)$, $h_f(s_2)$, $h_f(s_3)$, and $h_f(s_4)$ around $h_f(\Delta)$. Similarly, the order of s_1^* , s_2^* , s_3^* , and s_4^* around f agrees with the order of $h_f(s_1^*)$, $h_f(s_2^*)$, $h_f(s_3^*)$, and $h_f(s_4^*)$ around $h_f(f)$. By construction, the intersection of $h_f(l)$ and $h_f(l')$ in $h_f(f)$ is proper; hence the intersection of l and l' in f must also be proper. A similar argument holds if p lies on the boundary of a face f . Thus, by Lemma 2, every two arcs in $L(Q)$ intersect exactly once.

For every point $p \in \mathbf{P}^2$ and every $s \in l^*$, there is a unique pseudoline in $L(Q)$ passing through p and s . This pseudoline varies continuously as a function of s , sweeping over the whole projective plane as s runs through l^* . Hence, for any $p, q \in \mathbf{P}^2$, there is some pseudoline in $L(Q)$ containing p and q . If two distinct pseudolines $l, l' \in L(Q)$ both contained p and q , then l and l' would intersect more than once. Thus there must be a unique pseudoline $l(p, q) \in L(Q)$ containing p and q .

Finally, the continuity conditions on $l(p, q)$ and $l \cap l'$ follow from the fact that continuity is a local property, and that locally our pseudolines are nothing but homeomorphic images of lines.

We have thus proved

Theorem 1. *Every arrangement of pseudolines in the projective plane can be extended to a topological projective plane.*

We conclude this section with the proof of Lemma 1 promised above.

Proof of Lemma 1: Let \mathcal{A} be an arrangement of n pseudolines, at least $n - 1$ of which bound a face f of the arrangement, and let l be the n th pseudoline (or any one if all n bound f). l is stretchable, so there is some homeomorphism of the projective plane to itself that maps l to the line at infinity, l_∞ . Thus, without loss of generality, we may assume that $l = l_\infty$. Let $\{l_1, l_2, \dots, l_{n-1}\}$ be the remaining set of pseudolines, $\mathcal{A} \setminus \{l_\infty\}$.

If we remove l_∞ from the projective plane, we are left with a Euclidean plane, which we assume coordinatized. Each point $l_i \cap l_\infty$ can now be identified with some slope s_i in this plane. Without loss of generality, we can assume none of the slopes s_i is infinite. For each i , let l'_i be the line with slope s_i tangent to the unit circle, with l'_i chosen so that it passes above the unit circle if and only if l_i passes above f ; see Figure 5. (l_i passes above f if there exists some suitably large y_0 such that l_i separates f from $(0, y)$ for all $y > y_0$.)

We claim that the pseudoline arrangement $\{l_1, l_2, \dots, l_{n-1}\}$ is isomorphic to the straight line arrangement $\{l'_1, l'_2, \dots, l'_{n-1}\}$ in the Euclidean plane. Clearly this will imply that the arrangements $\{l_1, l_2, \dots, l_{n-1}, l_\infty\}$ and $\{l'_1, l'_2, \dots, l'_{n-1}, l_\infty\}$ are isomorphic in the projective plane.

The proof is by induction on the number of lines in the arrangements. It is trivially true for the arrangements $\{l_1\}$ and $\{l'_1\}$. Assume the arrangement $\mathcal{A}_k = \{l_1, l_2, \dots, l_k\}$ is isomorphic to the arrangement $\mathcal{A}'_k = \{l'_1, l'_2, \dots, l'_k\}$. Consider what happens when we add l_{k+1} and l'_{k+1} to \mathcal{A}_k and \mathcal{A}'_k , respectively. Without loss of generality, assume l_{k+1} lies below face f . Orient l_{k+1} so that face f lies to its left. Pseudoline l_{k+1} first intersects the pseudolines of \mathcal{A}_k lying above f whose slope is greater than s_{k+1} in order of increasing slope. l_{k+1} then intersects the pseudolines of \mathcal{A}_k lying below f in order of increasing slope. Finally, l_{k+1}

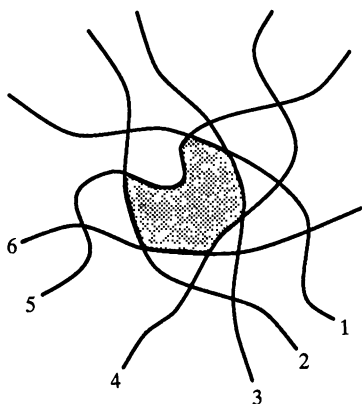


Figure 5a. Pseudolines around a face.

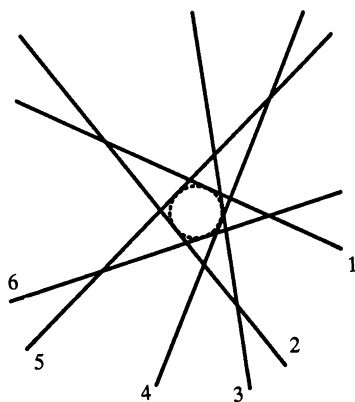


Figure 5b. Lines around the unit circle.

intersects the pseudolines of $-A_k$ lying above f whose slope is less than s_{k+1} in order of increasing slope. (If some pseudoline above f is “parallel” to l_{k+1} , we can consider it to be met by l_{k+1} after the rest.)

By construction, l'_{k+1} lies below the unit circle; orient it so that the unit circle lies to its left. Then l'_{k+1} also first intersects lines above the unit circle with slope greater than s_{k+1} , then lines below the unit circle, and finally the lines above the unit circle with slope less than s_{k+1} . The order in which l'_{k+1} meets them is also by increasing slope. Since l_i passes above f if and only if l'_i passes above the unit circle, and since the slopes of l_i and l'_i are equal, l_{k+1} and l'_{k+1} intersect corresponding pseudolines and lines in the same order. It follows that l_{k+1} and l'_{k+1} split corresponding faces in \mathcal{A}_k and \mathcal{A}'_k in exactly the same manner. Therefore the arrangements $\mathcal{A}_{k+1} = \{l_1, l_2, \dots, l_k, l_{k+1}\}$ and $\mathcal{A}'_{k+1} = \{l'_1, l'_2, \dots, l'_k, l'_{k+1}\}$ are isomorphic, and hence $\mathcal{A} \setminus \{l_\infty\}$ is isomorphic to the arrangement $\{l'_1, l'_2, \dots, l'_{n-1}\}$ of straight lines in the Euclidean plane. \square

3. UNIVERSAL TOPOLOGICAL PLANES. Recall that an arrangement \mathcal{A} of pseudolines is *stretchable* in Q if it is isomorphic to some arrangement of pseudolines $\mathcal{A}' \subset L(Q)$. In the previous section we proved that for every arrangement \mathcal{A} there is some topological projective plane Q in which \mathcal{A} is stretchable. We will now show that there is a topological plane in which *every* arrangement is stretchable.

We first need to introduce a technique for “patching” parts of one topological plane into another. Let Q and Q' be two topological planes. Let l_1, l_2 , and l_3 be pseudolines in Q that are not concurrent. l_1, l_2 , and l_3 then decompose Q into four closed regions which we call *triangles*. Let τ be any one of these triangles. The *vertices* of triangle τ are the points $l_1 \cap l_2, l_2 \cap l_3$, and $l_1 \cap l_3$; see Figure 6.

Similarly, choose three pseudolines in Q' that are not concurrent, label them l'_1, l'_2 , and l'_3 , and let triangle τ' be one of the closed regions bounded by these three pseudolines. Let ϕ be a homeomorphism of τ' onto τ that maps the vertices of τ' to the corresponding vertices of τ ; again, this will always exist by virtue of Schoenflies’ theorem [12].

Define a new topological plane Q'' , with a new set of pseudolines $L(Q'')$ chosen as follows. For each $l \in L(Q)$, if $l \cap \text{int}(\tau) = \emptyset$, then let l belong to $L(Q'')$. Otherwise, $l \cap \tau$ is a connected arc with two endpoints, p and q . Let l' be the

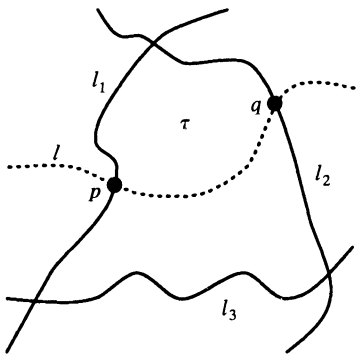


Figure 6a. Triangle τ .

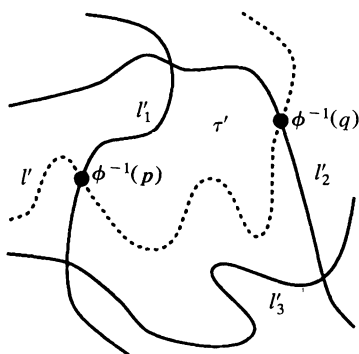


Figure 6b. Triangle τ' .

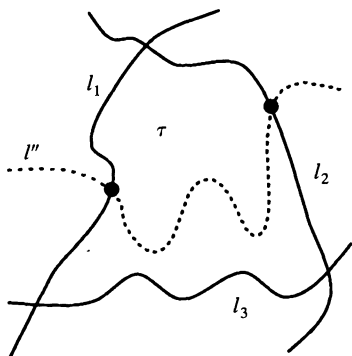


Figure 6c. Topological plane Q'' .

unique pseudoline in Q' passing through $\phi^{-1}(p)$ and $\phi^{-1}(q)$. Replace $l \cap \tau$ by $\phi(l' \cap \tau')$ to form a new pseudoline l'' , and let l'' belong to $L(Q'')$.

We must show that every two distinct pseudolines $l''_0, l''_1 \in L(Q'')$ intersect exactly once. But this is immediate from the fact that *outside* τ nothing has been altered, while *inside* τ two pseudolines meet if and only if their intersections with $\partial\tau$ interlace, a property which is preserved by the homeomorphism ϕ .

We can now prove

Theorem 2. *There exists a universal topological plane in which every arrangement of pseudolines is stretchable.*

Proof: Every arrangement of n pseudolines has at most $\binom{n}{2} + 1$ faces. Since the question of whether two arrangements are isomorphic depends only on the combinatorial structure of their associated cell complexes, there are only a finite number of isomorphism classes of arrangements of n pseudolines. Thus the set of isomorphism classes of arrangements of pseudolines of arbitrary size is countable.

Let $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3, \dots$ be a sequence of pseudoline arrangements such that every arrangement is isomorphic to some \mathcal{A}_i . Let Q_0 be any topological projective plane (for example \mathbf{P}^2), and let $\tau_1, \tau_2, \tau_3, \dots$ be a sequence of pairwise disjoint triangles in Q_0 . For each $i > 0$, construct the topological plane Q_i from the topological

plane Q_{i-1} as follows. Embed \mathcal{A}_i in some topological plane Q' , using Theorem 1. Let τ' be some triangle in Q' containing all the intersection points of the pseudolines in \mathcal{A}_i . (The existence of such a triangle τ' follows by continuity, since we can always find a pseudoline avoiding all the intersection points in \mathcal{A}_i and then take two other pseudolines sufficiently close to the first so that all the intersection points remain within a single region.) Replace triangle τ_i in Q_{i-1} by triangle τ' using the patching technique described above, to form the topological plane Q_i .

Let Q_∞ be the topological plane which is the limit of the topological planes Q_i . The pseudolines $L(Q_\infty)$ are formed from the original pseudolines in Q_0 by a (possibly infinite) sequence of local replacements. If two pseudolines $l, l' \in L(Q_\infty)$ intersected at two distinct points, p and q , then some corresponding pseudolines l_i and l'_i in some topological plane Q_i would also intersect at p and q , which is impossible. Thus Q_∞ is a topological plane. (The continuity conditions again follow as above, since they hold locally.)

Since all the intersections of pseudolines in \mathcal{A}_i occurred inside τ' , \mathcal{A}_i is isomorphic to some arrangement $\mathcal{A}'_i \subset L(Q_i)$. Because the triangle τ_i is never modified after the i th stage of the construction, it follows that \mathcal{A}_i is isomorphic to some arrangement in every Q_j , $j \geq i$, including Q_∞ . Thus every arrangement is stretchable in Q_∞ , i.e., Q_∞ is a universal topological plane. \square

Even though any universal topological plane contains all pseudoline arrangements, up to isomorphism, not all universal topological planes are isomorphic. In fact, using the techniques above, different choices of triangles and patchings may be shown to lead to uncountably many non-isomorphic universal planes [3], answering another question posed in [6].

4. OTHER DIRECTIONS. If we start with an arrangement of straight lines in the projective plane and let h_f be the identity map for every face f , then our construction will generate the standard projective plane \mathbf{P}^2 , with $L(\mathbf{P}^2)$ consisting of the usual straight lines. If h_f is not the identity, however, then even starting with a straight line arrangement we can generate a topological plane that is not isomorphic to the usual one.

In some sense, however, our construction does not generate topological planes whose pseudolines are too different from those in our original arrangement. An arrangement is called *k-piecewise linear* if each pseudoline in the arrangement is the union of at most k line segments. For any k , it turns out that there are pseudoline arrangements that are not isomorphic to any k -piecewise linear arrangement. (This can be shown, for example, by bounding the number of k -piecewise linear arrangements of n pseudolines using the Milnor-Thom theorem [9, 18] on the Betti numbers of solution sets of polynomial inequalities of a semi-algebraic set, and using the lower bound proved in [2] on the number of isomorphism classes of arrangements of n pseudolines.) On the other hand, if we start with an arrangement of n pseudolines and construct a topological plane Q by our methods, then any arrangement $\mathcal{A} \subset L(Q)$ can be shown to be isomorphic to some k -piecewise linear arrangement, where k depends only on n , the number of pseudolines in the original arrangement \mathcal{A} ; this fact plays an essential role in the proof of [3].

In [6], Grünbaum discusses an object intermediate between an arrangement of pseudolines and a topological plane. A *spread* of pseudolines is a 1-parameter family of pseudolines, any two meeting once, with the property that every point

$s \in l_\infty$ has a unique pseudoline through it which varies continuously as a function of s . In [6], Grünbaum asked if every arrangement can be extended to a spread, a question that we answered affirmatively in [4]. The much stronger result proved in Theorem 1 above is easily seen to imply that proposition. But our methods do not give any insight into the possibility of extending a spread to a topological plane, and this seems to be an intriguing question.

Finally, arrangements of pseudolines can be generalized to arrangements of pseudoplanes in dimension 3 (or of pseudohyperplanes in arbitrary dimension for that matter). These pseudoplanes should be “nicely” embedded in \mathbf{P}^3 , every two should intersect in a pseudoline of each, and every three should intersect in a single point. Can any arrangement of pseudoplanes be embedded in a continuous 3-parameter family of pseudoplanes, some sort of topological 3-space analogous to the topological space of planes in \mathbf{P}^3 ?

Just as we demanded of the points and pseudolines of a topological projective plane, the points, pseudolines and pseudoplanes of a topological projective 3-space should obey the standard incidence axioms of geometry, and their intersections should vary continuously. Desargues’ theorem, however, now turns out to be a direct consequence of these conditions, instead of being an independent axiom as it was in the plane [7]. Since any such geometry in which Desargues’ theorem holds is isomorphic to the standard one, any such topological projective 3-space turns out to be isomorphic to the usual \mathbf{P}^3 . On the other hand, there are certainly arrangements of pseudoplanes that are not isomorphic to arrangements of ordinary planes. This shows that there is no straightforward generalization of Theorem 1 possible to arrangements of pseudoplanes.

ACKNOWLEDGMENTS. We would like to express our appreciation to the Mittag-Leffler Institute, in Djursholm, Sweden, for its hospitality during the fall of 1991, for the opportunity it gave us to work together, and in particular to Anders Björner, Peter Mani, Jürgen Richter-Gebert, and Günter Ziegler, for many stimulating conversations.

REFERENCES

1. A. Björner, M. Las Vergnas, B. Sturmfels, N. White, and G. Ziegler, *Oriented Matroids*. Cambridge University Press, Cambridge, 1993.
2. J. E. Goodman and R. Pollack, Semispaces of configurations, cell complexes of arrangements, *J. Combinatorial Theory, Ser. A* 37 (1984), 257–293.
3. J. E. Goodman, R. Pollack, and R. Wenger, There are uncountably many universal topological planes (manuscript).
4. J. E. Goodman, R. Pollack, R. Wenger and T. Zamfirescu, Every arrangement extends to a spread. *Combinatorica* (to appear).
5. J. E. Goodman, R. Pollack, R. Wenger, and T. Zamfirescu, There is a universal topological plane. In *Proc. of the Eighth Annual ACM Symposium on Computational Geometry* (1992).
6. B. Grünbaum, *Arrangements and Spreads*. Amer. Math. Soc. Providence, 1972.
7. D. Hilbert, *The Foundations of Geometry*, 2nd ed. Open Court, Chicago, 1910.
8. F. Levi, Die Teilung der projektiven Ebene durch Gerade oder Pseudogerade. *Ber. Math.-Phys. Kl. sächs. Akad. Wiss. Leipzig* 78 (1926), 256–267.
9. J. Milnor, On the Betti numbers of real varieties. *Proc. Amer. Math. Soc.* 15 (1964), 275–280.
10. N. E. Mnëv, On manifolds of combinatorial types of projective configurations and convex polyhedra. *Soviet Math. Dokl.* 32 (1985), 335–337.
11. N. E. Mnëv, The universality theorems on the classification problem of configuration varieties and convex polytope varieties. In *Topology and Geometry—Rokhlin Seminar*, Lecture Notes in Mathematics 1346. Springer-Verlag, Heidelberg, 1988, pp. 527–543.
12. E. Moise, *Geometric Topology in Dimensions 2 and 3*. Springer-Verlag, New York, 1977.

13. F. R. Moulton, A simple non-Desarguesian plane geometry. *Trans. Amer. Math. Soc.* 3 (1902), 192–195.
14. J. Richter-Gebert, *On the Realizability Problem for Combinatorial Geometries—Decision Methods*. Ph.D. dissertation, Technische Hochschule Darmstadt, Darmstadt, 1992.
15. G. Ringel, Teilungen der projectiven Ebene durch Geraden oder topologische Geraden. *Math. Z.* 64 (1956), 79–102.
16. H. R. Salzmann, Topological planes. *Adv. Math.* 2 (1968), 1–60.
17. P. W. Shor, Stretchability of pseudoline arrangements is NP-hard. In *Applied Geometry and Discrete Mathematics—The Victor Klee Festschrift*. Amer. Math. Soc., Providence, 1991, pp. 531–554.
18. R. Thom, Sur l'homologie des variétés algébriques réelles. In *Differential and Combinatorial Topology*. Princeton University Press, Princeton, 1965, pp. 255–265.

Goodman:

*Department of Mathematics
City College, CUNY
New York, NY 10031
jegcc@cunyvm.cuny.edu*

Pollack:

*Courant Institute
New York University
New York, NY 10012
pollack@geometry.nyu.edu*

Wenger:

*Department of Computer Science
Ohio State University
Columbus, OH 43210
wenger@cis.ohio-state.edu*

Zamfirescu:

*Fachbereich Mathematik
Universität Dortmund
44221 Dortmund, GERMANY
zamfi@steinitz.mathematik.uni-dortmund.de*

Misunderstanding

Ah, you are a mathematician,
they say with admiration
or scorn.

Then, they say,
I could use you
to balance my checkbook.

I think about checkbooks.
Once' in a while
I balance mine,
just like sometimes
I dust high shelves.

From *Intersections: Poems by JoAnne Growney*,
Kadet Press, Bloomsburg, PA, 1993, p. 50.

An Application of Fourier Series to the Most Significant Digit Problem

Jeff Boyle

It is an old observation [N] that in tables of logarithms, the first few pages are more smudged and worn than the later pages. It follows that the distribution of the first significant digit of the numbers being looked up in the table is not uniform, but skewed toward the smaller digits. In fact, many sets of naturally arising numbers exhibit this asymmetry, with approximately 30% of the numbers beginning with the number 1 and less than 5% beginning with the number 9. Their first significant digit follows a logarithmic distribution, namely, the fraction of numbers whose first significant digit is n is approximately $\log(n + 1) - \log n$, $n = 1, 2, \dots, 9$.

This phenomenon is known as Benford's Law after Frank Benford, one of the first people to call attention to it. Indeed, he compiled a variety of data totalling 20,299 observations ranging from populations of cities to mathematical tables of powers of integers. The first significant digit in many (but not all) of his data sets were roughly logarithmically distributed.

Since Benford's Law is an empirically observed law of nature, rather than a theorem of pure mathematics, there are several competing explanations for it. This phenomenon of leading digits has been extensively investigated and enjoys a colorful history. For an interesting survey of this problem and its literature, see the expository article in the Monthly, August-September, 1976 by Ralph A. Raimi [R].

In all that follows we will write a positive number x in scientific notation, $x = m_x \cdot 10^{n_x}$, where $1 \leq m_x < 10$ and $n_x \in \mathbb{Z}$. For convenience we will refer to m_x as the mantissa of x . Of course, the most significant digit of x is $\llbracket m_x \rrbracket$, so it is sufficient to understand the distribution of m_x . One explanation of Benford's Law is that it results from "central limit-like" theorems for the mantissas of random variables under multiplicative operations. For example, the figures below show the density functions for the mantissa of the product of 1, 2, 3 and 4 independent and uniformly distributed random variables on $[1, 10]$. The density functions move progressively closer to the log distribution's density function $f(t) = 1/(t \ln 10)$.

Frequently, naturally-occurring data can be thought of as a result of products or quotients of random variables. For example, the population of a city tends to change roughly at a rate proportional to the population. If a city has an initial population P_0 and grows by $r_i\%$ in year i , then the population in n years is $P_n = P_0(1 + r_1)(1 + r_2) \cdots (1 + r_n)$, a product of a number of random variables. In this case, as we'll see, a set of populations of cities would follow Benford's law.

More specifically, we'll show for mantissas:

- (i) the log distribution is the limiting distribution when random variables are repeatedly multiplied, divided, or raised to integer powers,
- (ii) once achieved, the log distribution persists under all further multiplications, divisions, and raising to integer powers.

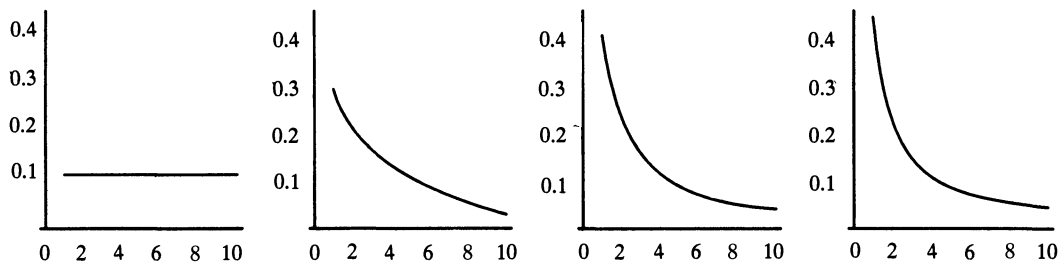


Figure 1. The p.d.f. for the mantissa of the products of 1, 2, 3, and 4 independent and uniformly distributed random variables on $[1, 10]$.

These facts have been explained to varying degrees by many other authors. Our purpose here is to show how these ideas can be neatly explained using Fourier Series.

The first step is to make the connection between mantissas under multiplication and modular arithmetic on $[0, 1]$ via logarithms. If random variables $x = m_x \cdot 10^{n_x}$ and $y = m_y \cdot 10^{n_y}$ are multiplied, the mantissa m_{xy} of the product is

$$m_{xy} = \begin{cases} m_x m_y & \text{if } 1 \leq m_x m_y < 10 \\ m_x m_y / 10 & \text{if } 10 \leq m_x m_y < 100. \end{cases}$$

This relationship is more easily studied logarithmically:

$$\log(m_{xy}) = \log(m_x) + \log(m_y) \pmod{1}. \tag{1}$$

Similarly, for division and raising to a power n ,

$$\log(m_{x/y}) = \log(m_x) - \log(m_y) \pmod{1}, \tag{2}$$

$$\log(m_{x^n}) = n \cdot \log m_x \pmod{1}. \tag{3}$$

An illuminating way to think about what is really going on is via the circular slide rule. The slide rule multiplies numbers (mantissas) by mechanically adding their logarithms ($2 \cdot 3 = 10^{.301} \cdot 10^{.477} = 10^{.788} = 6$). This can only work for all

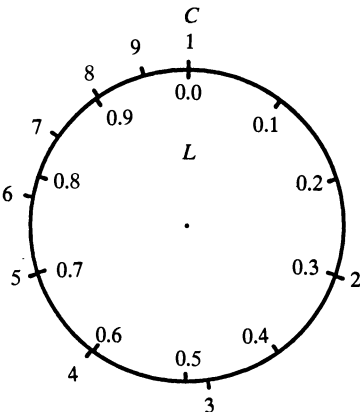


Figure 2. Numbers are not uniformly spaced around the c -scale but their logarithms on the L -scale are uniformly spaced.

multiplications if the logarithms are uniformly spread around the slide rule wheel. One can easily show that m_z has a log distribution $F(m) = P(m_z \leq m) = \log m$, if and only if $\log m_z$ has a uniform distribution on $[0, 1]$.

Now think of the circular slide rule as a roulette wheel. When we successively multiply random variables we are giving the “roulette wheel” a succession of random spins. No matter how unfairly we try to spin the wheel (assuming successive spins are independent), after a number of such spins are combined, the position of the wheel (as a random variable) is approximately uniformly distributed.

This approach to the log distribution has been widely studied. Largely overlooked, William Feller in his textbook, *An Introduction to Probability Theory and its Applications* [F], shows that under the most general conditions, the uniform distribution is the limiting distribution for sums of random variables taken mod 1. His method does not use Fourier Series and gives no rate of convergence. Other authors [T], [AS] have shown the log distribution is the limiting distribution for the mantissa of products for particular initial distributions, usually a uniform distribution on $[1, 10]$. See also [H] and [BB]. As far as I know, the Fourier Series approach was first applied by Furry and Hurwitz [FH] and more thoroughly exploited by Schatte [S1]. See also [M]. The approach we will use is essentially the same as [S1] although we generalize his results to include divisions and exponentiations. Although not necessary, for ease of exposition, we will assume our random variables are continuous on $[0, 1]$ and have a density function. In light of the earlier discussion, the following theorem implies that very generally, the log distribution is the limiting distribution for the first significant digit when continuous random variables are repeatedly multiplied.

Theorem 1. *Let x_1, \dots, x_k be independent and identically distributed continuous random variables on $[0, 1]$ with common density function $f \in L_2[0, 1]$, and let $z = \sum_{i=1}^k x_i \pmod{1}$. Then as $k \rightarrow \infty$, z approaches a uniform distribution.*

Actually, the above theorem is true under quite general circumstances. All that is required of the independent and identically distributed random variables x_i is that they are not discrete with their probability concentrated on a finite cyclic subgroup of $[0, 1] \pmod{1}$. (See [F], [S1], or [M]). These authors also show that if the x_i 's do have their probability concentrated on a finite cyclic subgroup of $[0, 1] \pmod{1}$, with $\text{prob}(x_i = 0) > 0$, then $\sum_{i=1}^k x_i$ approaches a discrete uniform distribution.

Since we are doing mod 1 arithmetic, it will be convenient to assume the density functions are defined on all of \mathbf{R} and are periodic of period 1. Hence if $f(t)$ is such a density

$$\int_a^{a+1} f(t) dt = 1 \quad \text{for all } a \in \mathbf{R}.$$

The Fourier coefficients of $f(t)$ are defined by

$$\hat{f}(n) = \int_0^1 f(t) \cdot e^{-2\pi i n t} dt, \quad n \in \mathbb{Z}. \quad (4)$$

Since $f(t)$ is real valued, $\hat{f}(-n) = \overline{\hat{f}(n)}$. If $f \in L^2[0, 1]$, then the partial sums of $\sum_{n=-\infty}^{\infty} \hat{f}(n) e^{2\pi i n t}$ converge to f in the L^2 norm. This is good enough for computing probabilities since L^2 convergence on $[0, 1]$ implies L^1 convergence. In fact, an application of Hölder's Inequality implies $\|f\|_1 \leq \|f\|_2$ on $[0, 1]$. The mapping $\mathcal{F}: L^2[0, 1] \rightarrow l^2$ given by $f(t) \rightarrow \{\hat{f}(n)\}_{n=-\infty}^{\infty}$ is a Hilbert space isomor-

phism, and consequently we have:

Parseval's Formula

$$\int_0^1 f^2(t) dt = \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2.$$

Lemma 1. *If $f \in L^2[0, 1]$ is a probability density, then,*

$$(a) \hat{f}(0) = 1 \quad (b) |\hat{f}(n)| < 1, \quad n \neq 0 \quad (c) \sup_{n \neq 0} |\hat{f}(n)| < 1.$$

Proof: (a) Since f is a density, $\hat{f}(0) = \int_0^1 f(t) \cdot e^0 dt = 1$.

$$(b) |\hat{f}(n)| = \left| \int_0^1 f(t) e^{-2\pi i n t} dt \right| \leq \int_0^1 f(t) |e^{-2\pi i n t}| dt = 1.$$

Equality could only hold if all the probability were concentrated on a finite cyclic subgroup of $[0, 1] \bmod 1$. This is not possible since we are only considering continuous random variables.

(c) As a consequence of Parseval's Formula, $|\hat{f}(n)| \rightarrow 0$ as $n \rightarrow \infty$. Combining this with (b) yields (c). ■

Now let x and y be continuous random variables on $[0, 1]$ with densities $f, g \in L^2[0, 1]$, respectively. If $z = x + y \pmod{1}$ and has density h , then

$$h(t) = \int_0^1 f(u) g(t - u) du = f * g, \text{ the convolution of } f \text{ and } g.$$

There is a nice equation relating the Fourier coefficients of f, g , and h .

Lemma 2. $\hat{h}(n) = \hat{f}(n) \cdot \hat{g}(n)$

Proof: Using the fact that both g and $e^{-2\pi i n t}$ are periodic of period 1, we have

$$\begin{aligned} \hat{h}(n) &= \int_0^1 h(t) \cdot e^{-2\pi i n t} dt = \int_0^1 \int_0^1 f(u) g(t - u) du \cdot e^{-2\pi i n t} dt \\ &= \int_0^1 \int_0^1 f(u) g(t - u) e^{-2\pi i n(t-u)} \cdot e^{-2\pi i n u} du dt \\ &= \int_0^1 f(u) e^{-2\pi i n u} \left(\int_0^1 g(t - u) e^{-2\pi i n(t-u)} dt \right) du = \hat{f}(n) \hat{g}(n). \end{aligned} \quad \blacksquare$$

Proof of Theorem 1: Recall x_1, \dots, x_k are independent and identically distributed continuous random variables on $[0, 1]$ with common density $f \in L^2[0, 1]$, and $z = \sum_{i=1}^k x_i \pmod{1}$. If z has density h , then by Lemma 2, $\hat{h}(n) = \hat{f}(n)^k$. As remarked earlier, to show z approaches a uniform distribution on $[0, 1]$ it suffices to show $h \rightarrow 1$ in the L^2 -norm. Thus

$$\begin{aligned} \|h - 1\|_2^2 &= \int_0^1 (h(t) - 1)^2 dt = \sum_{n \neq 0} |\hat{h}(n)|^2 = \sum_{n \neq 0} |\hat{f}(n)|^{2k} \\ &= \sum_{n \neq 0} |\hat{f}(n)|^2 |\hat{f}(n)|^{2k-2} \leq \left(\sup_{n \neq 0} |\hat{f}(n)| \right)^{2k-2} \cdot \sum_{n \neq 0} |\hat{f}(n)|^2. \end{aligned}$$

Since

$$\sup_{n \neq 0} |\hat{f}(n)| < 1 \quad \text{and} \quad \sum_{n \neq 0} |\hat{f}(n)|^2 < \infty,$$

we see that $\|h - 1\|_2^2 \rightarrow 0$ as $k \rightarrow \infty$. ■

Before generalizing Theorem 1 we would like to mention a couple of other striking consequences of Lemma 2. There exist many non-uniformly distributed pairs of random variables on $[0, 1]$ whose sum mod 1 is exactly uniformly distributed (or equivalently, there exist many pairs of random variables whose first significant digits are not logarithmically distributed, but the first significant digit of the product does have the log distribution). If x and y have densities $f, g \in L^2[0, 1]$ with $\hat{f}(n) \cdot \hat{g}(n) = 0$ for all $n \neq 0$, then $z = x + y \pmod{1}$ is uniformly distributed. For example, if $f(t) = 1 + \sin(2\pi t)$ and $g(t) = 1 + \sin(4\pi t)$, then z is uniform.

This leads to another observation, the invariance of the log distribution for first significant digits of products. Suppose x, y, z, f, g , and h are as in Lemma 2. Since $|\hat{g}(n)| < 1$ for all $n \neq 0$, $|\hat{h}(n)| \leq |\hat{f}(n)|$. If equality does not hold for some n , then $\|h - 1\|_2 < \|f - 1\|_2$, and so z is strictly closer to being uniform than x . This implies that the convergence in Theorem 1 is “monotone”. On the other hand, if equality holds for all n , then $\hat{f}(n) = 0 = \hat{h}(n)$ for all $n \neq 0$, and $f(t) = 1 = h(t)$, i.e., x and z are uniform. Therefore if $z = \sum_{i=1}^k x_i \pmod{1}$, and at least one of the x_i ’s is uniformly distributed, then z is uniformly distributed. This can be proved directly without using Fourier coefficients. See [H] for example. There is an amusing discrete analog to this result. Suppose one tosses a pair of dice, one of which is fair, and the other is of unknown nature. What can be said about the sum mod 6? Regardless of the nature of the second die the sum mod 6 must have uniform probability on $0, 1, \dots, 5$!

Next we would like to generalize Theorem 1 so that it applies to quotients and powers of random variables. Recall the connection between division and exponentiation with modular arithmetic on $[0, 1]$ via logarithms (see (2) and (3)). We need an equation relating the Fourier coefficients of the densities of a random variable $x \in [0, 1]$ and $z = mx \pmod{1}$, $m \in \mathbb{Z}$. The proof of the following lemma is left to the reader.

Lemma 3. *Let x be a continuous random variable on $[0, 1]$ with density $f \in L^2[0, 1]$, and let $z = mx \pmod{1}$, $m \in \mathbb{Z}$. If h is the density for z , then*

$$(a) \quad \hat{h}(n) = \hat{f}(mn) \quad \text{for all } n \in \mathbb{Z} \quad (b) \quad \text{if } m = -1, \hat{h}(n) = \hat{f}(-n) = \overline{\hat{f}(n)}.$$

Theorem 2. *Let x_1, \dots, x_k be independent and identically distributed continuous random variables on $[0, 1]$ with common density $f \in L^2[0, 1]$, and let $z = x_1 + \dots + x_j - x_{j+1} - \dots - x_k \pmod{1}$. Then z approaches a uniform distribution on $[0, 1]$ as $k \rightarrow \infty$.*

Proof: Let h be the density for z . Using Lemmas 2 and 3 we have

$$\begin{aligned} \|h(t) - 1\|_2^2 &= \sum_{n \neq 0} |\hat{h}(n)|^2 = \sum_{n \neq 0} |\hat{f}(n)|^{2j} \cdot \overline{|\hat{f}(n)|}^{2k-2j} \\ &= \sum_{n \neq 0} |\hat{f}(n)|^{2k} \rightarrow 0 \text{ as } k \rightarrow \infty. \end{aligned}$$

(See the proof of Theorem 1). ■

Theorem 3. Let x be a continuous random variable on $[0, 1]$ with density $f \in L^2[0, 1]$, and let $z = mx(\text{mod } 1)$, $m \in \mathbb{Z}$. Then z approaches a uniform distribution on $[0, 1]$ as $|m| \rightarrow \infty$.

Proof: Let h be the density of z . Then by Lemma 3

$$\|h(t) - 1\|_2^2 = \sum_{n \neq 0} |\hat{h}(n)|^2 = \sum_{n \neq 0} |\hat{f}(mn)|^2 \leq \sum_{|n| \geq m} |\hat{f}(n)|^2$$

This last sum goes to zero as $|m| \rightarrow \infty$ since $\sum_n |f(n)|^2$ converges. \blacksquare

Theorems 2 and 3 imply that the log distribution for first significant digits is the limiting distribution when continuous random variables are repeatedly multiplied, divided, or raised to integer powers. Let x_1, \dots, x_k be continuous random variables and let $n_j \in \mathbb{Z}$, $j = 1, \dots, k$. Then under quite general conditions on the x_i , if $z = x_1^{n_1} \cdots x_k^{n_k} / (x_1^{n_{k+1}} \cdots x_k^{n_k})$ is written as $z = m_z \cdot 10^{n_z}$, where $1 \leq m_z < 10$ and $n_z \in \mathbb{Z}$, then m_z approaches the log distribution as $\sum |n_i| \rightarrow \infty$.

One of the advantages of using Fourier Series is we can compute practical bounds on the rate of convergence to the log distribution. Several authors ([T], [AS], [A]) have investigated the convergence to the log distribution when random variables uniformly distributed on $[0, 1]$ or $[1, 10]$ are repeatedly multiplied. If X is such a random variable then one can easily show that $x = \log X(\text{mod } 1)$ has an exponential density function $f(t) = (\ln 10 \cdot 10^t / 9)$, $0 \leq t \leq 1$. Computing Fourier coefficients

$$\hat{f}(n) = \int_0^1 \frac{\ln 10 \cdot 10^t}{9} \cdot e^{-2\pi i n t} dt = \frac{\ln 10}{\ln 10 - 2\pi i n}.$$

Let x_1, \dots, x_k be independent random variables distributed as x above, and let $z = \sum_{i=1}^k x_i(\text{mod } 1)$. If z has density function h_k then by Lemma 2, $\hat{h}_k(n) = (\ln 10 / (\ln 10 - 2\pi i n))^k$. Hence,

$$\begin{aligned} \|h_k(t) - 1\|_2^2 &= \sum_{n \neq 0} \left| \frac{\ln 10}{\ln 10 - 2\pi i n} \right|^{2k} \leq 2 \cdot \sum_{n=1}^{\infty} \left(\frac{\ln 10}{2\pi n} \right)^{2k} \\ &= 2 \left(\frac{\ln 10}{2\pi} \right)^{2k} \sum_{n=1}^{\infty} \left(\frac{1}{n^{2k}} \right) \leq 2 \left(\frac{\ln 10}{2\pi} \right)^{2k} \left(\frac{2k}{2k-1} \right). \end{aligned}$$

Note that the L^1 distance between the densities of random variables x and y is the same as the L^1 distance between the densities of $g(x)$ and $g(y)$ for any one-to-one differentiable transformation g . In our situation, we are using the log transformation to convert random variables that are multiplied into ones that are added on $[0, 1]$. Since the L^2 norm on $[0, 1]$ dominates the L^1 norm, the L^2 norm provides a useful method for bounding the rate of convergence to the log distribution under repeated multiplications, divisions, or exponentiations. Therefore, Table 1(a) provides an accurate measure of the rate of convergence to the log distribution when k independent and uniformly distributed random variables on $[1, 10]$ are multiplied. For example, if five independent and uniformly distributed random variables on $[0, 1]$ are multiplied, the probability that the first significant digit is a one is $\log 2 \pm .0098$, or between .2912 and .3108.

In [AS] it is shown that if X is uniformly distributed on $[0, 1]$, then the first significant digit of X^m , $m \in \mathbb{Z}^+$ approaches a log distribution as $m \rightarrow \infty$. Of course, this is equivalent to showing that if the random variable x has density

TABLE 1. Convergence to the log distribution when uniformly distributed random variables on $[1, 10]$ are (a) multiplied, and (b) raised to integer powers.

k	bound on $\ h_k - 1\ _2$	m	bound on $\ h_m - 1\ _2$
1	.73	1	.66
2	.219	2	.33
3	.076	3	.22
4	.027	4	.17
5	.0098	5	.13
10	.000063	8	.083
∞	0	10	.066
		100	.0066
		1000	.00066
		∞	0

(a)
(b)

$f(t) = \ln 10 \cdot 10^t/9, 0 \leq t \leq 1$, then $x = mx(\text{mod } 1)$ approaches a uniform distribution. If $h_m(t)$ is the density for z , then by Lemma 3, $\hat{h}_m(n) = \hat{f}(mn) = \ln 10/(\ln 10 - 2\pi imn)$. Consequently,

$$\begin{aligned} \|h_m(t) - 1\|_2^2 &= \sum_{n \neq 0} \left| \frac{\ln 10}{\ln 10 - 2\pi imn} \right|^2 \leq \sum_{n \neq 0} \left| \frac{\ln 10}{2\pi imn} \right|^2 \\ &= 2 \left(\frac{\ln 10}{2\pi m} \right)^2 \sum_{n=1}^{\infty} \frac{1}{n^2} = 2 \left(\frac{\ln 10}{2\pi m} \right)^2 \frac{\pi^2}{6} = \frac{1}{12} \left(\frac{\ln 10}{m} \right)^2 \end{aligned}$$

Table 1(b) illustrates the convergence to the log distribution in this case.

Benford’s logarithmic law of first significant digits is very similar to another empirical law of nature, namely the common occurrence of bell shaped distributions. These naturally occurring distributions are a consequence of underlying mathematical forces. Certainly, as we have seen, the multiplicative operations—multiplication, division, and raising to powers—are a major contributing force in the log law. In fact, Benford log law is a consequence of “central limit” like theorems for first significant digits under the multiplicative operations.

ACKNOWLEDGMENTS. Finally, I would like to thank Bruce Riley, Mohammad Rahbar and Helen Skala for their help in writing this article. I learned about the first significant digit problem in a seminar taught by John Scheidt and Charles Schelin. I would like to thank them as well.

REFERENCES

[B] Benford, The law of anomalous numbers, *Proc. Amer. Phil. Soc.*, 78 (1938) 551–572.
 [N] Simon Newcomb, Note on the frequency of the use of the digits in natural numbers, *Amer. J. Math.* 4 (1881), 39–40.
 [R] Ralph A. Raimi, The First Digit Problem, *Amer. Math. Monthly* 83 (1976) 521–538.
 [F] William Feller, *An Introduction to Probability Theory and Its Applications*, Volume II, 1966 (Second edition) John Wiley and Sons, Inc., Page 273–4.
 [T] Peter R. Turner, The Distribution of Leading Significant Digits, *IMA Journal of Numerical Analysis* (1982) 2, 407–412.
 [AS] A. K. Adhikara, and B. D. Sarkar, Distributions of Most Significant Digit in Certain Functions Whose Arguments are Random Variables, *Sankhya: The Indian Journal of Statistics: Series B*, No. 30 (1968), 47–58.
 [FH] W. H. Furry, and H. Hurwitz, Distributions of numbers and distributions of significant figures, *Nature*, 155 (1945) 52–3.

- [S1] Peter Schatte, On the asymptotic uniform distributions of sums reduced mod 1, *Math. Nachr.* 115 (1984), 275–281.
- [S2] Peter Schatte, On sums modulo 2π of independent random variables, *Math. Nachr.* 110 (1983), 243–262.
- [M] K. V. Mardia, Statistics of Directional Data, 87–93.
- [A] A. K. Adhikari, Some results on the distribution of the most significant digit, *Sankhya: The Indian Journal of Statistics: Series B*. No. 31 (1969), 413–420.
- [H] R. W. Hamming, On the distribution of numbers, *The Bell System Technical Journal*, Vol. 49, No. 8, (1970).
- [BB] J. L. Barlow, and E. H. Bareiss, On the error distributions in floating point and logarithmic arithmetic, *Computing* 34, 325–347.

Mathematics Department
University of Wisconsin-La Crosse
La Crosse, WI 54601
boyle@math.uwlax.edu

Triangle Inequality

In a recent article, Wang and Zhang [2] prove that

$$[1 - |(u, v)|^2]^{1/2} \leq [1 - |(u, w)|^2]^{1/2} + [1 - |(v, w)|^2]^{1/2}$$

for unit vectors in a complex inner product space. As they observe, it is enough to prove this result when u, v, w are all in \mathbb{C}^2 . The theorem extends to all nonzero vectors in \mathbb{C}^2 if we define

$$d^2(u, v) = 1 - \frac{|(u, v)|^2}{|u|^2|v|^2};$$

we then have $d(u, v) \leq d(u, w) + d(w, v)$. A triangle inequality like this automatically attracts attention, and I want to analyze what lies behind it.

It is easy to see that $d(\lambda u, \mu v) = d(u, v)$ for nonzero complex scalars λ, μ . Thus if we map $u = (u_1, u_2)$ in $\mathbb{C}^2 \setminus \{(0, 0)\}$ to $z_u = u_2/u_1$ in $\mathbb{C} \cup \{\infty\}$, the value of d depends only on the images. Specifically, we can compute that

$$(*) \quad d(u, v) = d(z_u, z_v) = \frac{|z_u - z_v|}{[(1 + |z_u|^2)(1 + |z_v|^2)]^{1/2}}.$$

Now there is a standard bijection between the unit sphere and $\mathbb{C} \cup \{\infty\}$, given by projection from $(0, 0, 1)$ to the plane through the origin; and a straightforward computation [1, p. 22] shows that the distance between points on the sphere corresponding to z_u and z_v is exactly 2 times (*). Thus the triangle inequality in [2] is a consequence of the triangle inequality for distances on the Riemann sphere.

- [1] L. V. Ahlfors, Complex Analysis. McGraw-Hill, New York, 1953.
- [2] B.-Y. Wang and F. Zhang, A trace of inequality for unitary matrices. *Amer. Math Monthly* 101 (1994), 453–455.

William C. Waterhouse
Department of Mathematics
The Pennsylvania State University
University Park, PA 16802

NOTES

Edited by: John Duncan

Cross Product Identities in Arbitrary Dimension

Andrew Dittmer

I. INTRODUCTION. Recently, a certain amount of attention in the Monthly [1], [3] has been devoted to proofs of the vector identities on \mathbf{R}^3

- (1) (the “Baccab” identity) $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = \mathbf{b}(\mathbf{a} \cdot \mathbf{c}) - \mathbf{c}(\mathbf{a} \cdot \mathbf{b})$,
- (2) (usually obtained from (1)) $(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c})$.

It has been shown [2], [4] that the ordinary cross product generalizes to a function $\mathbf{a}_1 \times \cdots \times \mathbf{a}_{n-1}$ of $n - 1$ vectors on \mathbf{R}^n , and proofs were indicated in [2], [3] that (1) and (2) generalize to identities which hold in arbitrary dimension. In particular, we have

$$\begin{aligned} & \mathbf{a}_2 \times \cdots \times \mathbf{a}_{n-1} \times (\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}) \\ &= (-1)^{n+1} \begin{vmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_{n-1} \\ \mathbf{a}_2 \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_2 \cdot \mathbf{b}_{n-1} \\ \vdots & \cdots & \vdots \\ \mathbf{a}_{n-1} \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_{n-1} \cdot \mathbf{b}_{n-1} \end{vmatrix} \quad (1^*) \\ & (\mathbf{a}_1 \times \cdots \times \mathbf{a}_{n-1}) \cdot (\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}) = \begin{vmatrix} \mathbf{a}_1 \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_1 \cdot \mathbf{b}_{n-1} \\ \vdots & \cdots & \vdots \\ \mathbf{a}_{n-1} \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_{n-1} \cdot \mathbf{b}_{n-1} \end{vmatrix}, \quad (2^*) \end{aligned}$$

where in (1*) the determinant is to be interpreted formally in the manner of elementary calculus texts.

While several reasonably elementary proofs of the identities (1) and (2) are given in [1] and [3], only one ([3], p. 624, pp. 627–629) is coordinate-free and seems to generalize to higher dimensions. However, this argument contains a logical flaw which cannot trivially be patched in more than three dimensions. Since this line of reasoning is fairly common, we briefly review it here. Clearly, both sides of (1) are vectors orthogonal to \mathbf{a} and lie in the plane spanned by \mathbf{b} and \mathbf{c} . The claim is then made that this implies that the r.h.s. of (1) is a scalar multiple of the l.h.s.. The problem with this argument arises when \mathbf{a} is perpendicular to the plane spanned by \mathbf{b} and \mathbf{c} . If this occurs, any vector in this plane satisfies the condition of being orthogonal to \mathbf{a} , so we cannot trivially deduce that the r.h.s. is a scalar multiple of the l.h.s.. Of course, in this case, both sides are clearly zero. In the analogous

situation in higher dimensions, both sides of (1*) are zero. However, a proof of this seems to require an elaborate algebraic digression, which is alien from the geometric appeal of the original argument.

Perhaps the most elegant of the proofs of (1*) and (2*) already in the literature is given in [3], p. 625. In this approach we note that with the inner product given by $(\mathbf{a}_1 \wedge \mathbf{a}_2 \wedge \cdots \wedge \mathbf{a}_{n-1}) \cdot (\mathbf{b}_1 \wedge \mathbf{b}_2 \wedge \cdots \wedge \mathbf{b}_{n-1}) = \det(\mathbf{a}_i \cdot \mathbf{b}_j)$, the Hodge star map $^*: \Lambda^{n-1}(\mathbf{R}^n) \rightarrow \Lambda^1(\mathbf{R}^n) \cong \mathbf{R}^n$ becomes an isometry. It is then possible to define $\mathbf{a}_1 \times \mathbf{a}_2 \times \cdots \times \mathbf{a}_{n-1} := ^*(\mathbf{a}_1 \wedge \mathbf{a}_2 \wedge \cdots \wedge \mathbf{a}_{n-1})$ and then to verify that the standard properties of the extended cross product hold without much trouble.

The purpose of this note is to quickly review the definition and basic properties of the extended cross product and then to give a coordinate-free proof of identities (1*) and (2*) which uses only elementary linear algebra. We note that this proof works perfectly well in $n = 3$, in which case we obtain simple and completely elementary proofs of identities (1) and (2) (which proofs do not require introduction of the extended cross product). In the appendix, we give another proof employing the tensor product.

II. DEFINITION. We follow Spivak [4] in defining the cross product on \mathbf{R}^n . Let $\varphi: \mathbf{R}^n \rightarrow \mathbf{R}$ be defined by $\varphi(\mathbf{a}_1) = \det(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ (the determinant of the matrix with $\mathbf{a}_1, \dots, \mathbf{a}_n$ as row vectors), where $\mathbf{a}_2, \dots, \mathbf{a}_n$ are fixed vectors in \mathbf{R}^n . Since the determinant is linear in each of its variables, it follows that φ is linear as well. Therefore φ is a linear functional on \mathbf{R}^n , so there exists a unique vector $\mathbf{u} \in \mathbf{R}^n$ such that $\varphi(\mathbf{a}_1) = \mathbf{u} \cdot \mathbf{a}_1$. We then define $\mathbf{a}_2 \times \cdots \times \mathbf{a}_n = \mathbf{u}$. In other words, the cross product is characterized by

$$\mathbf{a}_1 \cdot \mathbf{u} = \det(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) \quad \text{for all } \mathbf{a}_1 \in \mathbf{R}^n \Leftrightarrow \mathbf{u} = \mathbf{a}_2 \times \cdots \times \mathbf{a}_n. \quad (3)$$

Since the triple scalar product identity $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \det(\mathbf{a}, \mathbf{b}, \mathbf{c})$ is well-known to hold for the standard cross product on \mathbf{R}^3 , it follows by (3) that the cross product as defined above coincides with the standard cross product when $n = 3$.

III. BASIC PROPERTIES. The basic properties satisfied by the cross product are:

The cross product is skew-symmetric and linear in each variable. (4)

$\mathbf{a}_2 \times \cdots \times \mathbf{a}_n$ is orthogonal to each of $\mathbf{a}_2, \dots, \mathbf{a}_n$. (5)

$\mathbf{a}_2 \times \cdots \times \mathbf{a}_n = \mathbf{0} \Leftrightarrow \mathbf{a}_2, \dots, \mathbf{a}_n$ are linearly dependent. (6)

If $\mathbf{a}_2 \times \cdots \times \mathbf{a}_n \neq \mathbf{0}$ then the n vectors $\mathbf{a}_2 \times \cdots \times \mathbf{a}_n, \mathbf{a}_2, \dots, \mathbf{a}_n$ form a positively oriented parallelotope. (7)

$|\mathbf{a}_2 \times \cdots \times \mathbf{a}_n|$ is the $n - 1$ -dimensional volume of the parallelotope formed by the vectors $\mathbf{a}_2, \dots, \mathbf{a}_n$. (8)

(4) follows immediately from the bilinearity of the dot product and the multilinearity and skew-symmetry of the determinant function. If in (3) we take $\mathbf{a}_1 = \mathbf{a}_i$ ($i = 2 \cdots n$), we obtain $\mathbf{a}_i \cdot (\mathbf{a}_2 \times \cdots \times \mathbf{a}_n) = \det(\mathbf{a}_i, \mathbf{a}_2, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n) = 0$, which proves (5). To see (6), note that by (3), $\mathbf{a}_2 \times \cdots \times \mathbf{a}_n = \mathbf{0}$ if and only if for all $\mathbf{a}_1 \in \mathbf{R}^n$, $\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \cdots \times \mathbf{a}_n) = \det(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = 0$. Hence for all $\mathbf{a}_1 \in \mathbf{R}^n$, $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ are linearly dependent, which happens when and only when $\mathbf{a}_2, \dots, \mathbf{a}_n$ are linearly dependent. Now the n -dimensional oriented volume of the parallelotope Δ formed by the vectors $\mathbf{a}_2 \times \cdots \times \mathbf{a}_n, \mathbf{a}_2, \dots, \mathbf{a}_n$ is

$$\begin{aligned} \det(\mathbf{a}_2 \times \cdots \times \mathbf{a}_n, \mathbf{a}_2, \dots, \mathbf{a}_n) &= (\mathbf{a}_2 \times \cdots \times \mathbf{a}_n) \cdot (\mathbf{a}_2 \times \cdots \times \mathbf{a}_n) \\ &= |\mathbf{a}_2 \times \cdots \times \mathbf{a}_n|^2 > 0 \end{aligned} \quad (9)$$

which shows (7). If $\mathbf{a}_2 \times \cdots \times \mathbf{a}_n = \mathbf{0}$, then (8) follows from (6). Otherwise, looking at Δ geometrically, we see that the vector $\mathbf{a}_2 \times \cdots \times \mathbf{a}_n$ is perpendicular to the hyperplane spanned by the other vectors. Therefore, the n -dimensional volume of Δ is Γh where h is the “height” $|\mathbf{a}_2 \times \cdots \times \mathbf{a}_n|$ and Γ is the “base” parallelotope formed by the vectors $\mathbf{a}_2, \dots, \mathbf{a}_n$. Comparison with (9) then gives (8).

IV. THE PROOF. Clearly, (1*) holds if and only if for all $\mathbf{a}_1 \in \mathbf{R}^n$,

$$\begin{aligned} & \mathbf{a}_1 \cdot [\mathbf{a}_2 \times \cdots \times \mathbf{a}_{n-1} \times (\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1})] \\ &= (-1)^{n+1} \begin{vmatrix} \mathbf{a}_1 \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_1 \cdot \mathbf{b}_{n-1} \\ \mathbf{a}_2 \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_2 \cdot \mathbf{b}_{n-1} \\ \vdots & \cdots & \vdots \\ \mathbf{a}_{n-1} \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_{n-1} \cdot \mathbf{b}_{n-1} \end{vmatrix} \end{aligned} \quad (10)$$

However, since an n -cycle is an even permutation if and only if n is odd,

$$\begin{aligned} & \mathbf{a}_1 \cdot [\mathbf{a}_2 \times \cdots \times \mathbf{a}_{n-1} \times (\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1})] \\ &= \det(\mathbf{a}_1, \dots, \mathbf{a}_{n-1}, \mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}) \\ &= (-1)^{n+1} \det(\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}, \mathbf{a}_1, \dots, \mathbf{a}_{n-1}) \\ &= (-1)^{n+1} (\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}) \cdot (\mathbf{a}_1 \times \cdots \times \mathbf{a}_{n-1}) \end{aligned} \quad (11)$$

follows from (3) and the skew-symmetry of the determinant. Therefore equations (1*) and (2*) are equivalent. In both this proof and the proof in the appendix we will show that (2*) holds.

If $\mathbf{b}_1, \dots, \mathbf{b}_{n-1}$ are linearly dependent, the l.h.s. of (2*) is zero by (6) and the r.h.s. is zero because the column vectors are linearly dependent. Otherwise, (6) implies that $|\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}| \neq 0$. In this case, define $n \times n$ matrices

$$M = \begin{pmatrix} \mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1} \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_{n-1} \end{pmatrix}, \quad N = (\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1} \quad \mathbf{b}_1 \quad \cdots \quad \mathbf{b}_{n-1}), \quad (12)$$

so M is a matrix of row vectors and N is a matrix of column vectors. By (3), we have

$$\det(M) = (\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}) \cdot (\mathbf{a}_1 \times \cdots \times \mathbf{a}_{n-1}), \quad \text{and} \quad (13)$$

$$\begin{aligned} \det(N) &= \det(N^T) = (\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}) \cdot (\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}) \\ &= |\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}|^2. \end{aligned} \quad (14)$$

Because of (5),

$$MN = \begin{pmatrix} |\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}|^2 & 0 & \cdots & 0 \\ \mathbf{a}_1 \cdot (\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}) & \mathbf{a}_1 \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_1 \cdot \mathbf{b}_{n-1} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{a}_{n-1} \cdot (\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}) & \mathbf{a}_{n-1} \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_{n-1} \cdot \mathbf{b}_{n-1} \end{pmatrix}. \quad (15)$$

The determinant of the submatrix

$$X = (\mathbf{a}_i \cdot \mathbf{b}_j) = \begin{pmatrix} \mathbf{a}_1 \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_1 \cdot \mathbf{b}_{n-1} \\ \vdots & \cdots & \vdots \\ \mathbf{a}_{n-1} \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_{n-1} \cdot \mathbf{b}_{n-1} \end{pmatrix} \quad (16)$$

is the r.h.s. of (2*). Since

$$\begin{aligned} & |\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}|^2 \det X \\ &= \det(MN) = \det(M)\det(N) \\ &= ((\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}) \cdot (\mathbf{a}_1 \times \cdots \times \mathbf{a}_{n-1})) |\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}|^2. \end{aligned} \quad (17)$$

and $|\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}| \neq 0$, the proof of equation (2*) is complete.

V. APPENDIX. The proof given here is a modified and generalized version of a proof given in [3] (see p. 624 and pp. 627–629). We first prove a lemma which shows that when a multilinear scalar-valued function is lifted to a function on the tensor product, the kernel of the resulting function can be determined solely by reference to the kernel of the original function.

Lemma 1. *Let V and W be finite-dimensional vector spaces over a field F . Let $p: V \times W \rightarrow F$ be a bilinear function, and let $p^*: V \otimes W \rightarrow F$ be the natural lift of p to a linear transformation of $V \otimes W$ guaranteed by the universal mapping property. Then $\ker(p^*)$ is spanned by the simple tensors $v \otimes w$ contained in it.*

Proof: For each $v \in V$, define a linear transformation $\varphi_v: W \rightarrow F$ by $\varphi_v(w) = p(v, w)$. Then the map $\varphi: v \mapsto \varphi_v$ defines a linear transformation from V to W^* , the dual space of W . By Gauss-Jordan elimination, we may choose bases $\{\mathbf{e}_i\}$ of V and $\{\mathbf{f}_i^*\}$ of W^* so that $\varphi(\mathbf{e}_i) = \mathbf{f}_i^*$ ($i = 1 \cdots k$), $\varphi(\mathbf{e}_i) = \mathbf{0}$ ($i > k$), where $k = \text{rank}(\varphi)$. If we now pass to the dual basis $\{\mathbf{f}_i\}$ of W , we have $p(\mathbf{e}_i, \mathbf{f}_j) = 1$ ($i = j \leq k$), $p(\mathbf{e}_i, \mathbf{f}_j) = 0$ otherwise. Now if p^* is identically zero, then the lemma is trivial, so we may assume that $\text{range}(p^*) = F$. If $\dim(V) = n$, $\dim(W) = m$, it follows that $\dim(\ker(p^*)) = nm - 1$, so it suffices to find $nm - 1$ linearly independent simple tensors in $\ker(p^*)$. However, if we set

$$S = \{\mathbf{e}_i \otimes \mathbf{f}_j: p^*(\mathbf{e}_i \otimes \mathbf{f}_j) = p(\mathbf{e}_i, \mathbf{f}_j) = 0\}, \quad (18)$$

$$T = \{\mathbf{e}_1 \otimes \mathbf{e}_1 - \mathbf{e}_1 \otimes \mathbf{e}_i + \mathbf{e}_i \otimes \mathbf{e}_1 - \mathbf{e}_i \otimes \mathbf{e}_i = (\mathbf{e}_1 + \mathbf{e}_i) \otimes (\mathbf{e}_1 - \mathbf{e}_i): 1 < i \leq k\}, \quad (19)$$

then S has $nm - k$ elements, T has $k - 1$ elements, and $S \cup T$ is clearly linearly independent and contained in $\ker(p^*)$.

The following corollary results from a straightforward induction on r .

Corollary. *Let V_1, \dots, V_r be finite-dimensional vector spaces over a field F . Let $p: V_1 \times \cdots \times V_r \rightarrow F$ be a multilinear function, and let $p^*: V_1 \otimes \cdots \otimes V_r \rightarrow F$ be the natural lift of p to a linear transformation of $V_1 \otimes \cdots \otimes V_r$. Then $\ker(p^*)$ is spanned by the simple tensors $v_1 \otimes \cdots \otimes v_r$ contained in it.*

The next lemma is crucial in this proof of (2*). An elementary proof of a weaker statement (which is, however, sufficient for proving the identity under consideration) can be found in [3].

Lemma 2. Let $p, q: V_1 \times \cdots \times V_r \rightarrow F$ be multilinear functions, and suppose that $\ker(p) \subseteq \ker(q)$. Then there exists a scalar λ in F such that $p\lambda = q$.

Proof: Lift p, q to linear transformations $p^*, q^*: V_1 \otimes \cdots \otimes V_r \rightarrow F$. Now $\ker(p) \subseteq \ker(q)$ implies that the set of simple tensors in $\ker(p^*)$ is contained in the set of simple tensors in $\ker(q^*)$. However, the preceding corollary shows that these tensors span all of $\ker(p^*)$ and $\ker(q^*)$ respectively, so $\ker(p^*) \subseteq \ker(q^*)$. If q^* is identically zero, then the lemma is trivial, so assume otherwise. Then $\text{range}(p^*) = \text{range}(q^*) = F$, so

$$\dim(\ker(p^*)) = \dim(\ker(q^*)) = \dim(V_1 \otimes \cdots \otimes V_r) - 1.$$

Setting $K = \ker(p^*) = \ker(q^*)$ there is a tensor $\mathbf{t} \notin K$ so that $V_1 \otimes \cdots \otimes V_r = K + \langle \mathbf{t} \rangle$. Then $\lambda = q^*(\mathbf{t})/p^*(\mathbf{t})$ clearly satisfies the requirements.

Let $V_i = \mathbf{R}^n$ ($i = 1, \dots, 2n - 2$), and interpret each side of (2*) as a multilinear function in $\mathbf{a}_1, \dots, \mathbf{a}_{n-1}, \mathbf{b}_1, \dots, \mathbf{b}_{n-1}$. If we could show that in this case the hypotheses of Lemma 2 are satisfied, it would show that the r.h.s. of (2*) is a constant multiple of the l.h.s. Evaluating for $\mathbf{a}_i = \mathbf{b}_i = \mathbf{e}_{i+1}$ ($\{\mathbf{e}_i\}$ denoting the standard orthonormal basis of \mathbf{R}^n) would then show that this constant is 1, giving us a proof of (2*).

Proving that the hypotheses of Lemma 2 hold for the sides of (2*) is clearly equivalent to showing that the r.h.s. of (2*) is zero whenever the l.h.s. is zero. In the trivial cases where $\mathbf{a}_1, \dots, \mathbf{a}_{n-1}$ (or $\mathbf{b}_1, \dots, \mathbf{b}_{n-1}$) are linearly dependent, the l.h.s. of (2*) is zero by (6) and the r.h.s. is zero because the row vectors (column vectors) are linearly dependent. Otherwise, the fact that

$$\begin{aligned} 0 &= (\mathbf{a}_1 \times \cdots \times \mathbf{a}_{n-1}) \cdot (\mathbf{b}_1 \times \cdots \times \mathbf{b}_{n-1}) \\ &= \det(\mathbf{a}_1 \times \cdots \times \mathbf{a}_{n-1}, \mathbf{b}_1, \dots, \mathbf{b}_{n-1}) \end{aligned} \quad (20)$$

implies that the vector $\mathbf{a}_1 \times \cdots \times \mathbf{a}_{n-1}$ is a nontrivial linear combination of the vectors $\mathbf{b}_1, \dots, \mathbf{b}_{n-1}$. Therefore, the columns of the r.h.s. of (2*) nontrivially generate the vector

$$(\mathbf{a}_1 \cdot (\mathbf{a}_1 \times \cdots \times \mathbf{a}_{n-1}), \dots, \mathbf{a}_{n-1} \cdot (\mathbf{a}_1 \times \cdots \times \mathbf{a}_{n-1})) = \mathbf{0}, \quad (21)$$

which shows that the r.h.s. of (2*) is zero whenever the l.h.s. is.

ACKNOWLEDGMENTS. The proof of Lemma 1 appearing in this paper is due to William Pardon.

REFERENCES

1. Paul Binding, More on cross products, this Monthly, 98 (1991) 850–851.
2. R. Shaw, Vector cross products in n dimensions, *Int. J. Math. Educ. Sci. Technol.*, 18 (1987) 803–816.
3. R. Shaw and F. J. Yeadon, On $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$, this Monthly, 96 (1989) 623–629.
4. M. Spivak, *Calculus on Manifolds*, W. A. Benjamin, Inc., Menlo Park, California, 1965, pp. 83–84.

Department of Mathematics
Duke University
Durham, NC 27708

A Non-Constant Continuous Function on the Plane Whose Integral on Every Line Is Zero

D. H. Armitage

If f is a continuous function on \mathbb{C} , integrable on \mathbb{C} with respect to plane Lebesgue measure, and if

$$\int_l f \, ds = 0 \tag{1}$$

for every (doubly infinite) line l , where s denotes length measure, then f must be identically zero. This has been known for many years (see [5] for references), but only comparatively recently was it shown that this result fails in the absence of the global integrability condition. In fact, Zalcman [5; pp. 243–244], using a theorem of Arakelian [1; p. 1189] concerning tangential holomorphic approximation on unbounded sets, constructed a non-constant entire function f satisfying (1) on every line l . The purpose of this note is to give a construction, similar to Zalcman's, but using only elementary complex analysis. As a bonus, all the derivatives of the function we produce also have vanishing integrals on all lines.

The main idea is to construct a non-constant entire function g whose derivatives satisfy

$$\int_l |g^{(n+1)}| \, ds < +\infty \tag{2}$$

and

$$g^{(n)}(z) \rightarrow 0 \quad \text{as } z \rightarrow \infty \quad (z \in l) \tag{3}$$

for each line l and each non-negative integer n (cf. [5; p. 243]). If we define $f = g'$, then by (2) each $f^{(n)}$ is integrable on every line l , while (3) together with the fundamental theorem of calculus shows that

$$\int_l f^{(n)} \, ds = 0 \quad (n = 0, 1, 2, \dots)$$

for each line l .

We construct g by a simple pole-pushing technique which goes back at least to Runge [4].

Lemma. *Suppose that $z_1, z_2 \in \mathbb{C}$ and $|z_1 - z_2| < 1$. If ϕ_1 is holomorphic on $\mathbb{C} \setminus \{z_1\}$ and $\varepsilon > 0$, then there exists a holomorphic function ϕ_2 on $\mathbb{C} \setminus \{z_2\}$ such that*

$$|(\phi_2 - \phi_1)(z)| < \varepsilon(1 + |z|)^{-2} \quad (|z - z_2| > 1). \tag{4}$$

For the sake of completeness, we indicate a proof of this lemma. The function ϕ_1 has a Laurent expansion centred at z_2 :

$$\phi_1(z) = \phi_0(z) + \sum_{j=1}^{\infty} a_j(z - z_2)^{-j} \quad (|z - z_2| > |z_1 - z_2|),$$

where ϕ_0 is entire. By choosing m sufficiently large, we can arrange that the function ϕ_2 , defined by

$$\phi_2(z) = \phi_0(z) + \sum_{j=1}^m a_j(z - z_2)^{-j} \quad (z \neq z_2),$$

satisfies (4).

To complete the construction of g , choose a sequence (ζ_k) of points lying on the parabolic arc $P = \{t + it^2: t \geq 0\}$ such that

$$\zeta_0 = 0, \quad |\zeta_k - \zeta_{k-1}| < 1 \quad (k \geq 1), \quad \zeta_k \rightarrow \infty.$$

Let $g_0(z) = z^{-2}$. Using the lemma repeatedly, we obtain a sequence (g_k) of functions such that g_k is holomorphic on $\mathbb{C} \setminus \{\zeta_k\}$ and

$$|(g_k - g_{k-1})(z)| < 2^{-k}(1 + |z|)^{-2} \quad (k \geq 1, |z - \zeta_k| > 1). \quad (5)$$

Then (g_k) is locally uniformly convergent to a limit function g which is entire. Define $P_a = \{z: \inf_{w \in P} |z - w| > a\}$. If $z \in P_1$, then by (5)

$$|(g - g_0)(z)| \leq \sum_{k=1}^{\infty} |(g_k - g_{k-1})(z)| < (1 + |z|)^{-2} < |g_0(z)|,$$

so that $g \neq 0$ and

$$|g(z)| < (1 + |z|)^{-2} + |g_0(z)| < 2|z|^{-2} \quad (z \in P_1).$$

Hence, by Cauchy's estimates,

$$|g^{(n)}(z)| < 2n!|z|^{-2} \quad (n = 0, 1, 2, \dots, z \in P_2).$$

Since $l \setminus P_2$ is bounded for any line l , it follows that (2) and (3) hold, as required.

The real part of the function f constructed above (or of Zalcman's function) provides an example of a non-constant harmonic function on \mathbb{R}^2 whose integral on every line is zero. Recently, Armitage and Goldstein [3] showed that there exists a non-constant harmonic function h on \mathbb{R}^N , where $N \geq 2$, such that

$$\int_{\Lambda} h \, d\lambda = 0$$

for every $(N - 1)$ -dimensional hyperplane Λ , where λ denotes $(N - 1)$ -dimensional measure. The above argument for holomorphic functions can be mimicked with harmonic functions on \mathbb{R}^N ; for harmonic pole-pushing see, for example, [2]. Thus the construction in [3] can be made more elementary. In particular, the result of [3] can be obtained without recourse to the quite difficult harmonic approximation theorem [2; Theorem 1.1] that we used, which played a role analogous to that of Arakelian's theorem in [5].

REFERENCES

1. N. V. Arakelian, Uniform approximation on closed sets by entire functions, *Izv. Akad. Nauk SSSR Ser. Mat.*, 28 (1964), 1187–1206.
2. D. H. Armitage and M. Goldstein, Better than uniform approximation on closed sets by harmonic functions with singularities, *Proc. London Math. Soc.*, 60 (1990), 319–343.

3. D. H. Armitage and M. Goldstein, Nonuniqueness for the Radon transform, *Proc. Amer. Math. Soc.*, 117 (1993), 175–178.
4. C. Runge, Zur Theorie der eindeutigen analytischen Funktionen, *Acta Math.*, 6 (1885), 228–244.
5. L. Zalcman, Uniqueness and nonuniqueness for the Radon transform, *Bull. London Math. Soc.*, 14 (1982), 241–245.

Department of Pure Mathematics
Queen's University
Belfast BT7 1NN
Northern Ireland
d.armitage@uk.ac.qub.v2

Chu's 1303 Identity Implies Bombieri's 1990 Norm-Inequality (Via an Identity of Beauzamy and Dégot)

Doron Zeilberger¹

Blessed are the meek: for they shall inherit the earth (Matthew V.5)

Inequalities are deep, while *equalities* are shallow. Nevertheless, it sometimes happens that a deep inequality, **A**, follows from a mere *equality* **B**, which, in turn, follows from a more general, and *trivial*² identity **C**.

In this note we demonstrate this, following [3], with **A** := Bombieri's norm inequality [2]³, **B** := an identity of Reznick [5], and **C** := an identity of Beauzamy and Dégot [3]. This exposition differs from the original only in the punch line: I give a 1-line proof of **C**, using Chu's identity.

Let $P(x_1, \dots, x_n)$ and $Q(x_1, \dots, x_n)$ be two polynomials in n variables:

$$P = \sum_{i_1, \dots, i_n \geq 0} a_{i_1, \dots, i_n} x_1^{i_1} \cdots x_n^{i_n}, \quad Q = \sum_{i_1, \dots, i_n \geq 0} b_{i_1, \dots, i_n} x_1^{i_1} \cdots x_n^{i_n}.$$

The *Bombieri inner product* [2] is defined by

$$[P, Q] := \sum_{i_1, \dots, i_n \geq 0} (i_1! \cdots i_n!) \cdot a_{i_1, \dots, i_n} b_{i_1, \dots, i_n},$$

and the *Bombieri norm*, 'by: $\|P\| := \sqrt{[P, P]}$.

Bombieri's Inequality A. Let P and Q be any *homogeneous* polynomials in (x_1, \dots, x_n) , then

$$\|PQ\| \geq \|P\| \|Q\|.$$

¹This note was written while the author was on leave (Fall 1993) at the Institute for Advanced Study, Princeton. I would like to thank Don Knuth for a helpful suggestion.

²Trivial to verify, not to conceive!

³It was needed by Beauzamy and Enflo in their research on deep questions on Banach spaces. It also turned out to have far reaching applications to computer algebra! [1].

In order to state **B** and **C**, we need to introduce the following notation. $D_i := \partial/\partial x_i$, ($i = 1, \dots, n$), $P^{(i_1, \dots, i_n)} := D_1^{i_1} \cdots D_n^{i_n} P$, and for any polynomial $A(x_1, \dots, x_n)$, $A(D_1, \dots, D_n)$ denotes the linear partial differential operator with constant coefficients obtained by replacing x_i by D_i .

A follows almost immediately from ([5][3]):

Reznick's Identity B. For any polynomials P, Q in n variables:

$$\|PQ\|^2 = \sum_{i_1, \dots, i_n \geq 0} \frac{\|P^{(i_1, \dots, i_n)}(D_1, \dots, D_n)Q(x_1, \dots, x_n)\|^2}{i_1! \cdots i_n!}.$$

Beauzamy and Dégot's Identity C. For any polynomials P, Q, R, S in n variables:

$$\begin{aligned} & [PQ, RS] \\ &= \sum_{i_1, \dots, i_n \geq 0} \frac{[R^{(i_1, \dots, i_n)}(D_1, \dots, D_n)Q(x_1, \dots, x_n), P^{(i_1, \dots, i_n)}(D_1, \dots, D_n)S(x_1, \dots, x_n)]}{(i_1! \cdots i_n!)}. \end{aligned}$$

Proof of B \Rightarrow A. Pick the terms for which $i_1 + \cdots + i_n$ equals the (total) degree of P , let's call it p , and note that for those (i_1, \dots, i_n) , $P^{(i_1, \dots, i_n)}(x_1, \dots, x_n) = (i_1! \cdots i_n!)a_{i_1, \dots, i_n}$, so

$$\begin{aligned} & \sum_{i_1 + \cdots + i_n = p} \frac{\|P^{(i_1, \dots, i_n)}(D_1, \dots, D_n)Q(x_1, \dots, x_n)\|^2}{i_1! \cdots i_n!} \\ &= \sum_{i_1 + \cdots + i_n = p} \|a_{i_1, \dots, i_n}Q(x_1, \dots, x_n)\|^2 \cdot (i_1! \cdots i_n!) \\ &= \left(\sum_{i_1 + \cdots + i_n = p} (a_{i_1, \dots, i_n})^2 \cdot (i_1! \cdots i_n!) \right) \|Q(x_1, \dots, x_n)\|^2 = \|P\|^2 \|Q\|^2. \end{aligned}$$

Proof of C \Rightarrow B. Take $R = P$ and $S = Q$.

Proof of C. Both sides are linear in P , in Q , in R , and in S , so it suffices to take them all to be typical monomials, ($P = x_1^{p_1} \cdots x_n^{p_n}$, and similarly for Q, R , and S), for which the assertion follows immediately by applying Chu's [4] identity⁴

$$\sum_{i \geq 0} \binom{r}{i} \binom{s}{p-i} = \binom{r+s}{p},$$

to $r = r_t$, $s = s_t$, $p = p_t$, ($t = 1 \cdots n$), (using i_t for i), and taking their product.
Q.E.D.

REFERENCES

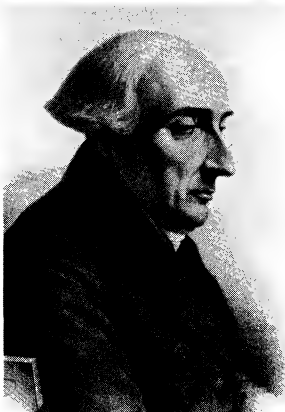
1. B. Beauzamy, Products of polynomials and a priori estimates for coefficients in polynomial decompositions: A sharp result, *J. Symbolic Computation* **13** (1992), 463–472.
2. B. Beauzamy, E. Bombieri, P. Enflo, and H. L. Montgomery, Products of polynomials in many variables, *J. Number Theory* **36** (1990), 219–245.

⁴Rediscovered in the 18th century by Vandermonde. Proved by counting, in two different ways, the number of ways of picking p lucky winners out of a set of r boys and s girls.

3. B. Beauzamy and J. Dégot, Differential Identities, I.C.M., Paris, *Proc. Amer. Math. Soc.*, to appear.
4. Chu Chi-kie, manuscript, 1303, China. (See J. Needham, *Science and Civilization in China*, v. 3, Cambridge University Press, New York, 1959.)
5. B. Reznick, An inequality for products of polynomials, *Proc. Amer. Math. Soc.* 117 (1993), 1063–1073.

*Department of Mathematics
Temple University
Philadelphia, PA 19122
zeilberg@math.temple.edu*

Who are these mathematicians and what did they share?



Answer on page 910.

THE COMPUTER SCIENCE SAMPLER

Edited by: Catherine C. McGeoch

How to Stay Competitive

Catherine C. McGeoch

In the Office Of The Future, the coffee pot will come to you. You'll push a button on your wrist communicator, and a mobile robot will appear at your desk with fresh pots of coffee and tea, and donuts. Also, through the Wonder of Technology, you will only have to work a half-day to maintain your current productivity, so you can devote the rest of your time to entrepreneurial pursuits.

Suppose you decide to go into the mobile-coffee-robot business, providing coffee-serving robots to office buildings for a flat monthly fee. Your expenses are proportional to the distance the robots travel, so you want to minimize this quantity by making smart decisions about which robot to send to service any given request. (After servicing a request, the robot stays where it is until needed again.) Of course it would help if you could make some predictions about future requests, but that is not possible. Here are two strategies:

Greedy. Always send the robot that is nearest to the request.

Balance. Send the robot S such that the total distance traveled by S (so far this month) would be minimized over all robots.

Which strategy is better? How can you even compare them if you don't know anything about the request sequence? It turns out that by a certain reasonable standard described below, Greedy is a bad algorithm; and in some cases Balance is a good algorithm. This column analyses the two algorithms according to this standard.

The k -server problem. Let \mathcal{G} be the set of symmetric weighted graphs on n nodes having nonnegative edge weights (the nodes represent offices and the weights represent distances between offices). There are k servers that move from node to node by traversing edges. The cost of traversing an edge is equal to the weight on that edge. The minimum cost of moving a server from node w to node x (not necessarily via a direct edge) is d_{wx} .

A sequence $\sigma = \langle \sigma_1, \sigma_2 \dots \sigma_m \rangle$ of m requests for service at vertices in $G \in \mathcal{G}$ is presented. Suppose σ_i is a request at a particular vertex x . A k -server algorithm A must *cover* node x by moving one of the k servers there. (We assume throughout that $2 \leq k < n$ since otherwise all algorithms are identical.) If A moves a server from node w then the *service cost* is d_{wx} . We assume that if there is already a server on node x , then the algorithm does nothing and the service cost

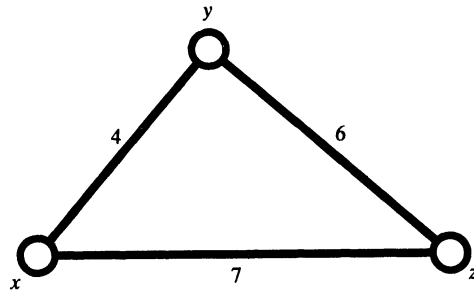


Figure 1

is 0. Let $C_A(\sigma_t)$ denote the service cost for algorithm A to cover request σ_t , and let $C_A(\sigma) = \sum_{t=1}^m C_A(\sigma_t)$ be the total cost of A on sequence σ .

In fact the cost of A on σ also depends on the initial placement of the servers on nodes (before the first request). This technical point will be addressed in the proofs below, which hold for arbitrary initial configurations.

An algorithm is *online* if it must decide how to cover σ_t without having information about future requests. Both Greedy and Balance are online. An algorithm that can examine all of σ before deciding how to cover each request σ_t is *offline*.

An online k -server algorithm B is *c-competitive* if, for any graph $G \in \mathcal{G}$, any request sequence σ , and all other k -server algorithms A , it holds that $C_B(\sigma) \leq c \cdot C_A(\sigma) + a$, for a constant a that may depend on G . Note that A is allowed to be either online or offline.

If B is c -competitive, then your mobile-coffee-robots will travel no more than c times as far as is absolutely necessary to service requests. An algorithm for which c is a small constant is a good algorithm. How small can c be? It is known that there cannot be a c -competitive algorithm for the k -server problem when $c < k$ (see [3]). The two theorems below concern the k -competitiveness of Greedy and Balance.

Theorem. *Greedy is not k -competitive for the k -server problem.*

Proof: We need only exhibit a counterexample to k -competitiveness. Let $k = 2$, and consider the 3-node graph in Figure 1 and the sequence $\sigma = \langle z, x, y, x, y, x, y, x, \dots \rangle$ of length m . The first two requests will cause Greedy to position the two servers on nodes x and z (with cost between 0 and 6 depending on initial conditions). After that, Greedy will shuttle one server between nodes x and y , incurring a service cost of 4 each time. Therefore the total cost for Greedy is $C_G(\sigma) \geq 4(m - 2)$.

However an online algorithm A could place its servers on x and y after the first three requests (incurring cost at most 12), and subsequently have a service cost of 0. Therefore $C_A(\sigma) \leq 12$. For any constant a , the sequence σ can be made long enough to obtain $C_G(\sigma) \not\leq 2C_A(\sigma) + a$. It is straightforward to extend this argument to larger k by using this example graph as a subgraph in a larger problem. \square

This theorem can be extended to show that Greedy is not c -competitive for any constant c . Greedy can be “fooled” with a request sequence that makes it behave

badly in comparison to an offline algorithm. The next theorem shows that in the special case $k = n - 1$, Balance can't be similarly fooled: any sequence that is bad for Balance is also fairly bad (within a factor of k) for all algorithms. To get some intuition, try running Balance on the above example.

First we define the Balance strategy a little more carefully. Suppose σ_t is a request for vertex s . If s is already covered, then Balance does nothing. If s is not covered, then for each node l that is currently covered, let $D_{l,t-1}$ be the distance traveled so far by the server on l . Balance covers s by moving a server from some node w such that $D_{w,t-1} + d_{ws}$ is minimized over all nodes. After the move we have $D_{s,t} = D_{w,t-1} + d_{ws}$. The Lemma below states that under the Balance rule, the difference in distance traveled, between the most-recently-moved server and all other servers, is bounded above and below by constants. See [3] for a proof of this Lemma.

Lemma. *At any time t , let s be the most recently requested vertex, let w be the vertex left vacant by Balance, and let l be any other covered vertex. Then $-d_{sl} \leq D_{l,t} - D_{s,t} \leq d_{lw} - d_{ws}$.*

Theorem. *Balance is k -competitive for the k -server problem when $k = n - 1$.*

Proof: We must show that for any G there is a constant a such that the following bound holds:

$$\forall A, \forall \sigma, C_B(\sigma) \leq (n - 1)C_A(\sigma) + a$$

where A is an arbitrary algorithm for the k -server problem and B denotes the Balance algorithm.

Note that when $k = n - 1$, exactly one vertex is vacant at any time. We assume that the sequence σ is *hard* for Balance: that is, σ always forces Balance to move by always requesting the vacant vertex. If Balance is k -competitive on hard sequences then it is k -competitive on all sequences.

Now consider the situation just after σ_t is covered. Let x_t be the vertex left vacant by A and let y_t be the vertex left vacant by Balance. It suffices to exhibit a *residue function* $R(x_t, y_t, \sigma_t, t)$, with $R(x_0, y_0, \sigma_0, 0)$ defined to be 0, for which two properties hold. Property 1 is that for $0 \leq t < m$,

$$R(x_{t+1}, y_{t+1}, \sigma_{t+1}, t + 1) \leq R(x_t, y_t, \sigma_t, t) + (n - 1)C_A(\sigma_{t+1}) - C_B(\sigma_{t+1}).$$

Then

$$\begin{aligned} & \sum_{t=1}^m R(x_t, y_t, \sigma_t, t) - R(x_{t-1}, y_{t-1}, \sigma_{t-1}, t - 1) \\ & \leq \sum_{t=1}^m (n - 1)C_A(\sigma_t) - \sum_{t=1}^m C_B(\sigma_t), \end{aligned}$$

and therefore

$$R(x_m, y_m, \sigma_m, m) - R(x_0, y_0, \sigma_0, 0) \leq (n - 1)C_A(\sigma) - C_B(\sigma).$$

Property 2 is that there is a constant $-a$ (dependent on G) such that

$$-a \leq R(x_m, y_m, \sigma_m, m).$$

With these two properties established, a little algebraic manipulation completes the proof. The residue function represents a comparison of the service costs of

Balance and A , for arbitrary A . Properties 1 and 2 bound the residue function to ensure that the total costs of Balance and A cannot get too far apart.

We now define R . To simplify notation, for fixed t let $s = \sigma_t$, let $v = x_t$ be the vertex left vacant by A after covering s , and let $w = y_t$ be the vertex left vacant by Balance. After Balance moves, $D_{l,t}$ is the total distance traveled by the server currently on node l . Let $D_t^* = \sum_l D_{l,t}$ be the total over all servers on covered nodes. Then let

$$R(v, w, s, t) = \begin{cases} (n-1)(D_{s,t} - d_{ws}) - D_t^* & \text{if } v = w \\ (n-1)D_{v,t} - D_t^* & \text{if } v \neq w, s \end{cases}$$

Note that $R(s, w, s, t)$ is undefined since vertex s cannot be left vacant by A .

Now consider the next request σ_{t+1} . Since σ is assumed to be hard for Balance, we know that $\sigma_{t+1} = w$. Balance moves a server from some vertex k such that $D_{k,t} + d_{kw}$ is minimized. The service cost for Balance is $C_B(\sigma_{t+1}) = d_{kw}$. After covering w , A leaves some vertex j vacant. The service cost is $C_A(\sigma_{t+1})$, equal to either d_{jw} or 0, depending on whether w was previously covered by A .

To establish property 1, we must show that for all j, v where $R(j, k, w, t+1)$ and $R(v, w, s, t)$ are defined,

$$R(j, k, w, t+1) \leq R(v, w, s, t) + (n-1)C_A(\sigma_{t+1}) - C_B(\sigma_{t+1}).$$

We proceed by case analysis using the following observations.

1. After Balance covers w , $D_{w,t+1} = D_{k,t} + d_{kw}$. For all other covered nodes l , $D_{l,t+1} = D_{l,t}$. Therefore $D_{t+1}^* = D_t^* + d_{kw}$.
2. If $v = w$, then $R(v, w, s, t) = (n-1)(D_{s,t} - d_{ws}) - D_t^*$ and $C_A(\sigma_{t+1}) = d_{jw}$.
3. If $v \neq w$, then $R(v, w, s, t) = (n-1)D_{v,t} - D_t^*$. Since w is already covered by A , we know that $C_A(\sigma_{t+1}) = 0$. Since A doesn't move to cover w , we know that $j = v$.
4. If $j = k$, then $R(j, k, w, t+1) = (n-1)(D_{w,t+1} + d_{kw}) - D_{t+1}^*$. If $j \neq k$, then $R(j, k, w, t+1) = (n-1)D_{j,t+1} - D_{t+1}^*$. In either case, using observation 1 we obtain $R(j, k, w, t+1) = (n-1)D_{j,t} - D_t^* - d_{kw}$.

Case A. Suppose $v = w$. Then using observations 2 and 4, we must show that

$$(n-1)D_{j,t} - D_t^* - d_{kw} \leq (n-1)(D_{s,t} - d_{ws}) - D_t^* + (n-1)d_{jw} - d_{kw},$$

or, cancelling terms, that

$$D_{j,t} - D_{s,t} \leq d_{jw} - d_{ws}.$$

Application of the Lemma finishes the proof of Property 1 for this case.

Case B. Suppose $v \neq w$. Then using observations 3 and 4, we must show that

$$(n-1)D_{j,t} - D_t^* - d_{kw} \leq (n-1)D_{v,t} - D_t^* - d_{kw}.$$

Since $j = v$ here the result is immediate.

Property 2 remains. Inspection of the definition of $R(v, w, s, t)$ shows that the residue function has a constant lower bound for any t (including $t = m$) if, for any vertex i , $(n-1)D_{i,t} - D_t^*$ has a constant lower bound. This fact also follows from the Lemma. (Note that n is a constant here). \square

More about k servers. Variations on the k -server problem have many computational applications besides coffee-serving robots. For example, most computer systems maintain a large *secondary memory* (on disks) and a much smaller *main*

memory (on RAM chips). Programs and data can only be processed if they are residing in main memory. Only k pages (or chunks) of information can be in main memory at a given time. This can be viewed as a k -server problem: the nodes represent all n pages in the system, and the servers mark the k pages that are currently in main memory. An edge weight corresponds to the cost of moving (or “swapping”) a page between secondary and main memory; we usually assume that all weights are identical. A page request is determined by the currently running program, and the operating system must try to minimize the total number of swaps without knowing which pages will be needed in the future.

It can be shown that if $k \neq n - 1$ then Balance is not k -competitive. A 2-competitive 2-server algorithm does exist; otherwise, for $2 < k < n - 1$ it is an open problem to find a k -competitive algorithm for the k -server problem.

This problem was first described by Manasse, McGeoch, and Sleator [3] who proved all of the results mentioned in this column and conjectured that a k -competitive k -server algorithm must exist. The conjecture has prompted intensive research and led to several nice results about the nature of online computation. Recent attention has focused on an algorithm described by Chrobak and Larmore [1]; Koutsoupias and Papadimitriou [2] have shown that that algorithm is $(2k - 1)$ -competitive for the general k -server problem.

ACKNOWLEDGMENT. Thanks go to Lyle McGeoch for several stimulating discussions regarding the k -server problem.

REFERENCES

1. M. Chrobak and L. L. Larmore, The server problem and on-line games, in *On-Line Algorithms*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol 7, L. A. McGeoch and D. D. Sleator, editors, American Mathematical Society, 1992, pp. 11–64.
2. E. Koutsoupias and C. Papadimitriou, On the k -server conjecture, *Proceedings of the 26th ACM Symposium on Theory of Computing*, May (1994) 507–511.
3. M. Manasse, L. A. McGeoch, and D. D. Sleator, Competitive algorithms for server problems, *Journal of Algorithms* 11 (1990) 208–230.

*Department of Mathematics and Computer Science
Amherst College
Amherst, MA 01002
ccm@cs.amherst.edu*

Each generation has its few great mathematicians, and mathematics would not even notice the absence of the others. They are useful as teachers, and their research harms no one, but it is of no importance at all. A mathematician is great or he is nothing.

—Alfred Adler
“Mathematics and Creativity,”
The New Yorker Magazine,
February 19, 1972.

THE EVOLUTION OF . . .

Edited by Abe Shenitzer

Mathematics, York University, North York, Ontario M3J 1P3, Canada

On the Calculus of Variations and Its Major Influences on the Mathematics of the First Half of Our Century. Part II.*

Erwin Kreyszig

Note: The first part of this paper (sections 1–5) appeared in 1994, in the August–September issue of the *Monthly* (pp. 674–678). What follows is a short summary of the first part and the concluding part of the paper.

Summary of Sections 1–5. The calculus of variations deals with the problem of determination of extrema of functionals given by integrals. It may be said to have begun with Johann Bernoulli's brachistochrone problem of 1696. The birthyear of its *theory* is 1744, the year in which Euler published his *Methodus inveniendi* . . . that included his necessary condition for a minimum and a splendid collection of problems.

Lagrange went beyond Euler by inventing the “method of variations” and the concept of the first variation δJ of a functional J ; δJ is the analogue of the first derivative of a function.

Both Euler and Lagrange contributed to the problem of minimal surfaces, an important geometric application of the calculus of variations.

The discovery of sufficient conditions for an extremum of a functional was due to Jacobi and Weierstrass. Legendre introduced the concept of the second variation $\delta^2 J$ of a functional J , which is an analog of the second derivative of a function. Weierstrass rigorized the calculus of variations and introduced the fundamental concepts of a field of extremals and of the *E*-function, “a turning point in the history of the calculus of variations.”

6. IMPACT ON EARLY FUNCTIONAL ANALYSIS. The proverbial *Weierstrassian rigor* (Felix Klein's term of 1885) had a profound influence on the theories of *functionals* and *function spaces* of our century. In fact, near the end of the last century, the central role of functionals in the calculus of variations may very well have directed attention to functionals in general. This may have been a kind of subliminal influence that affected primarily the younger generation represented by Volterra and later by Fréchet and F. Riesz.

*Abbreviated version of a paper with the same title.

Five notes of 1887 by Volterra on special classes of functionals, investigated as concepts *per se*, marked the *birth of functional analysis*. Influenced by Dini and Betti, the latter a close friend of Riemann's, Volterra wanted to *generalize complex analysis*. His whole theory was based on the calculus of variations.

Hadamard was enthusiastic about Volterra's novel theory and contributed a basic result that attained final form in a celebrated paper by F. Riesz. (Riesz's theorem states that every continuous linear functional $U[f]$, $f \in C[a, b]$, can be expressed as a Stieltjes integral

$$U[f] = \int_a^b f(x) dw(x),$$

where $w(x)$ is determined by U and is of bounded variation on $[a, b]$.) Hadamard pointed to the variational-analytic roots of functional analysis—the *calculus* gave rise to the *calculus of variations*, which in turn produced the *functional calculus*, later called functional analysis.

Hadamard regarded the calculus of variations as

“a first chapter of functional calculus, whose development will without doubt be one of the first tasks in the analysis of the future.”

(This quotation is from Hadamard's book on the calculus of variations published in 1910.)

7. DIRICHLET'S PRINCIPLE. Besides its direct impact on developing functional analysis from 1887 to about 1903 the calculus of variations also had an indirect and, in the long run, an even greater impact on (classical and) functional analysis, an impact which dates back to about 1870 and reached functional analysis shortly after the turn of the century. This involved *partial differential equations*, for which the development proceeded from the search for general solutions to solution formulas for boundary and initial value problems (Green, Poisson, Kirchhoff, etc.) and on to existence (and uniqueness) proofs, notably for Laplace's equation for a function u of three variables

$$\Delta u = u_{xx} + u_{yy} + u_{zz} = 0 \quad (7.1)$$

(or in two variables x, y), which is basic in gravitation, electrostatics, stationary heat conduction, and fluid flow. For the corresponding *Dirichlet problem* in a general domain,

$$\Delta u = 0 \text{ in } \Omega; \quad u|_{\partial\Omega} = f, \quad u \in C^2(\Omega) \cap C^0(\bar{\Omega}), \quad \Omega \subset \mathbb{R}^2 \text{ or } \mathbb{R}^3, \quad (7.2)$$

a proof of existence of a solution u is not easy and attracted the efforts of many of the greatest mathematicians for quite some time. A (faulty) method of proof was provided by a principle taken from the *calculus of variations*. B. Riemann (1826–66) had first seen it in lectures of G. Lejeune Dirichlet (1805–59) in Berlin and named it after him:

Dirichlet's principle. There exists a function u that minimizes the functional (the so-called *Dirichlet integral*)

$$D[u] = \int_{\Omega} |\text{grad } u|^2 dV, \quad \Omega \subset \mathbb{R}^2 \text{ or } \mathbb{R}^3, \quad (7.3)$$

among all functions $u \in C^1(\Omega) \cap C^0(\bar{\Omega})$ which take on given values f on the boundary $\partial\Omega$ of Ω , and that function u satisfies (7.2).

Note that Laplace's equation is the Euler equation for (7.3).

We see that the Dirichlet integral is bounded below (by zero) and the claim of existence of a minimum is based on a conceptual error, namely the failure to distinguish *greatest lower bound* from *minimum*.

Dirichlet's principle was used earlier (in 1839) by C. F. Gauss (1777–1855) in his potential-theoretic investigations. Gauss claimed that if V is the potential of a mass distribution of density m on the surface S of a spatial region and U is a given function on S , then among all possible distributions there *obviously* exists one for which

$$\int (V - 2U) m dS$$

takes its minimum. If this is granted, then one can show, as Gauss did, that: (i) for this minimizing distribution, $V - U = \text{const}$ at all points of S that carry mass; (ii) if $U = 0$, there must be mass everywhere on S ; (iii) one can obtain a distribution whose potential V on S equals U . Thus one could conclude the existence of the required *harmonic function* [a twice *continuously* differentiable solution of (7.1)] as well as its representation as the potential of a single layer of mass.

After Gauss, Dirichlet's principle was used in 1847 by W. Thomson (Lord Kelvin, 1824–1907) in order to “prove” the existence of a solution u of the differential equation

$$(\alpha^2 u_x)_x + (\alpha^2 u_y)_y + (\alpha^2 u_z)_z = 4\pi\zeta \quad (7.4)$$

that vanishes at infinity. Here α and ζ are given functions and ζ is zero outside a given bounded region. Because of this work, the principle is usually called in England *Thomson's principle*.

After attending Dirichlet's lectures in Berlin, Riemann used the principle in his famous thesis of 1851 as a key tool for obtaining fundamental results on *complex* analytic functions from *real* potential theory. Since the principle does not always hold, some of Riemann's proofs were not complete. All the results, however, turned out to be correct and were proved later by other methods. The same holds for Riemann's later use of Dirichlet's principle in his monumental paper on the theory of Abelian functions.

It was a strange situation. Dirichlet's principle had helped to produce exciting basic results but doubts about its validity began to appear, first in private remarks of Weierstrass—which did not impress Riemann, who placed no decisive value on the derivation of his existence theorems by Dirichlet's principle—and then, after both Dirichlet and Riemann had died, in Weierstrass's public address to the Berlin Academy:

“From Dirichlet's assumptions it can only be claimed that for (7.3) there exists a certain lower bound to which (7.3) can come arbitrarily close, without being forced to actually reach it. Dirichlet's argument appears invalid.”

8. ANOTHER IMPACT ON FUNCTIONAL ANALYSIS. The breakdown of Dirichlet's principle had an enormous positive effect on analysis because it led to the creation of three ingenious new methods for obtaining existence proofs for the Dirichlet problem—by H. A. Schwarz (1843–1925), H. Poincaré (1854–1912), and C. Neumann (1832–1925)—as well as to the development of the direct methods of the calculus of variations initiated by D. Hilbert (1862–1943).

C. Neumann's *method of the arithmetic mean* (1870) was of great importance to analysis and to functional analysis because it sparked work on spectral theory, integral equations, and through it on Hilbert spaces. Neumann assumed the solution to be the potential of a *double layer*, a layer of dipoles normal to the boundary surface (or curve) $\partial\Omega$, of unknown density $\sigma(Q)$, $Q \in \partial\Omega$. Writing this potential in the form

$$u(P) = \frac{1}{2\pi} \int_{\partial\Omega} \sigma(Q) \frac{\partial}{\partial\nu} \left(\frac{1}{r} \right) dS(Q), \quad r = d(P; Q), \quad Q \in \partial\Omega, \quad (8.1)$$

(ν the outer normal of $\partial\Omega$) one obtains for the unknown density σ the integral equation

$$\sigma(Q) = \frac{1}{2\pi} \int_{\partial\Omega} \frac{\partial}{\partial\nu} \left(\frac{1}{r^*} \right) \sigma(Q^*) dS(Q^*) = \varphi(Q),$$

$$\varphi(Q) \text{ the given values of } u \text{ on } \partial\Omega, \quad r^* = d(Q, Q^*). \quad (8.2)$$

Because of the term $\sigma(Q)$, the operator form of this “integral equation of the second kind” is

$$(I + K)\sigma = \varphi. \quad (8.3)$$

Its solution should obviously be

$$\sigma = (I + K)^{-1} \varphi = \varphi - K\varphi + K^2\varphi - K^3\varphi + \cdots. \quad (8.4)$$

Using this idea, Neumann was able to prove existence of a solution of the Dirichlet problem by integral equation methods. He solved (8.2) by successive approximation, defining $\sigma_0 = \varphi$ and

$$\sigma_n = (-K)\sigma_{n-1} = -\frac{1}{2\pi} \int_{\partial\Omega} \sigma_{n-1} \frac{\partial}{\partial\nu} \left(\frac{1}{r} \right) dS = (-K)^n \varphi.$$

This gave him the *Neumann series* (8.4) as the solution of the problem in a convex domain in space or in the plane.

We conclude this part of the section with a short list of events resulting from Neumann's work. In 1888, Weierstrass's former student du Bois-Reymond coined the term *integral equations* and expressed the desirability of a general theory of these equations, with which one could solve various problems such as that solved by Neumann. It did not take long for such theories—by Le Roux (1895), Volterra (1896), and the most famous one by Fredholm (1900, 1903)—to appear. Hilbert “caught fire at once” (as H. Weyl put it) and developed his spectral theory of integral equations with symmetric kernel published in six *Mitteilungen* between 1904 and 1910. Most important of these was the fourth *Mitteilung*, the earliest truly functional-analytic treatment of integral equations, in which he introduced continuous and compact (Hilbert said “completely continuous”) forms (cast into operator language by F. Riesz in 1913).

After the breakdown of Dirichlet's principle there was no more *general* principle for handling various problems of applied mathematics. It seems that in that situation Hilbert first put his hope in the calculus of variations, which had produced general principles in the past. In Problem 23 of his famous talk of 1900 on unsolved problems Hilbert had drawn attention to Weierstrass's work and to A. Kneser's book, the first presentation of the modern calculus of variations. Not intimidated by Weierstrass, he was able to re-establish the Dirichlet principle within proper limits as a valid method of proof. He did this in two papers of 1900

and 1901 (reprinted 1905). In the first of these papers he proposed the following more general formulation of Dirichlet's principle.

“Every regular problem of the calculus of variations [Sec. 3] has a solution as soon as suitable restrictive assumptions with respect to the nature of the given boundary conditions are satisfied and, if necessary, the concept of a solution is suitably generalized.”

This approach gave rise to the *direct methods* (methods without the use of the Euler-Lagrange equations), which became of basic importance in the existence theory of the calculus of variations. (A forerunner of these methods was Euler's almost forgotten “direct difference method”.) Another solution of the Dirichlet problem by direct methods was given later (in 1907) by Lebesgue.

Apart from these splendid initial steps, Hilbert made no further attempts to uniformize analysis by methods of the calculus of variations. Instead, he turned to integral equations, perhaps as a more promising tool for the same purpose. But his idea of weakening the notion of solutions became a guiding principle in the calculus of variations of our century.

9. PLATEAU'S PROBLEM. By *Plateau's problem* one means the determination of a simply connected portion of a minimal surface S in R^3 bounded by a given curve in space. This problem is named after the Belgian physicist J. Plateau, who realized minimal surfaces experimentally by dipping wires (the boundary curves) into soap solutions, minimum area corresponding to minimum surface energy.

Plateau's problem has attracted great attention from around 1870 to the present. It is a problem genuinely belonging to the calculus of variations. The solution methods developed by a pleiad of researchers (Schwarz, Lebesgue, Korn, Bernstein, Haar, Garnier, Radó, Douglas, Courant, McShane) were interrelated with various branches of the mathematics of our century, which they fertilized immensely.

From a more general viewpoint we can regard the evolution of the theory of minimal surfaces and of Plateau's problem as particular cases of the development of the theory of partial differential equations. The first stage concerned special solutions of the minimal surface equation (special minimal surface), while the second stage dealt with general solutions (Weierstrass's general solution formulas). The third stage, the solution of boundary value problems (Plateau's problem), began in 1867 with Schwarz's work, slightly later than work on partial differential equations in general. For this work, Kelvin, Gauss, Riemann, and others had already switched from general solutions to the geometrically and physically more useful boundary and initial value problems.

10. GLOBAL CALCULUS OF VARIATIONS (MORSE THEORY). We have seen that early functional analysis owed much to the calculus of variations, and that *general topology* (*set-theoretic topology*) developed along with it, in a process of mutual give-and-take that extended over the first three decades of our century. This led to the creation of modern nonlinear analysis in connection with partial differential equations, and even more in connection with the calculus of variations, resulting in the so-called *calculus of variations in the large*, or *Morse theory*, for short.

In this calculus one is concerned with relations between properties of a “space” X (usually a topological space, often a Riemannian manifold) and a real-valued continuous function f defined on X .

The beginnings of this fascinating theory are due to Poincaré, who began his work on periodic solutions of the differential equations of celestial mechanics in his thesis of 1879 and knew of “Morse inequalities” for a surface as early as 1885, seven years before Morse was born.

The next stage of the development can perhaps best be seen from G. D. Birkhoff’s book on Dynamical Systems (1927). In it Birkhoff emphasized the growing importance of topology in the calculus of variations. He also mentioned his Ph.D. student M. Morse, who worked out his calculations in the large in his book published by the AMS in 1934. “In the large” meant that Morse considered the whole manifold on which the variational problem was given and not just in a small neighborhood of an extremal (“calculus of variations in the small”). In the Preface he commented:

“Any problem which is nonlinear in character, which involves more than one coordinate system or more than one variable, or whose structure is initially defined in the large, is likely to require consideration of topology . . .”

Around 1930 topology was developing rapidly and eliciting general interest. This was evidenced, for example, by the wealth of new results contained in Alexandroff-Hopf’s *Topologie I* of 1935 and by the papers presented at the Moscow Topology Congress of 1935. Thus it was just the right time for a marriage of the classical calculus of variations and topology, and Morse made ingenious use of the latter.

Define a *critical point* of a smooth function f on a smooth manifold M to be a point p at which

$$\text{grad } f = 0,$$

Morse classified the critical points in terms of the eigenvalues of the Hessian matrix

$$H(f, p) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j} \right]_p \quad (10.1)$$

and obtained *topological* lower bounds (in terms of Betti numbers) for the number of critical points. These are the famous *Morse inequalities*. This work applied to *functions*. From *functions*, Morse proceeded to *functionals*, essentially n -dimensional analogs of our integral (2.1) and generalizations. Now for *functions* the topology at that time was sufficient. For *functionals*, that is, functions on a space of curves, Morse had to develop a topology in function spaces. Instead of critical points he had *critical curves*. He described this intuitively in his talk of 1932 at the Zurich International Congress. From it, one gains the impression that in Morse’s theory the calculus of variations and topology had developed multiple relations. Incidentally, a similar theory was created simultaneously and independently by L. A. Lusternik and L. Schnirelman. This shows that at certain times certain things are *in the air*, in the sense that problems which extend known settings in a natural way become accessible as soon as basic theories (topology in the present case) have been sufficiently developed.

Let me conclude with the following remark. We started with the calculus, sketched briefly how calculus evolved into the calculus of variations, and outlined the most important ways in which the calculus of variations accompanied, influenced, or even initiated progress in various parts of analysis, geometry, functional analysis and, finally, topology. The whole process was heterogeneous, but I hope

that we have seen traces of some intrinsic logic of the development here and there. The presentation can perhaps help to bring about a better understanding of certain features of present-day mathematics. The calculus of variations seems to have had a profound effect on the general development of mathematics. I hope that readers will see this as an invitation to further research pertaining to details as well as to larger issues that call for a more profound study than that presented in these pages.

REFERENCES

1. G. F. Simmons, *Differential Equations with Applications and Historical Notes* (second edition), McGraw-Hill, 1991, 1972. See Chapter 12.
2. N. I. Akhiezer, *Calculus of Variations*, Blaisdell, 1962.
3. I. G. Petrovsky, *Lectures on Partial Differential Equations*, Dover reprint, 1992.
4. G. M. Ewing, *Calculus of Variations with Applications*, Norton, 1969.

Department of Mathematics and Statistics
Carleton University
Ottawa, Ontario
Canada, K1S 5B6

I do hate sums. There is no greater mistake than to call arithmetic an exact science. There are permutations and aberrations discernible to minds entirely noble like mine; subtle variations which ordinary accountants fail to discover; hidden laws of number which it requires a mind like mine to perceive. For instance, if you add a sum from the bottom up, and then from the top down, the result is always different.

—Mrs. La Touche
Mathematical Gazette, vol. 12.

THE AUTHORS

ROBERT GRAY studied mathematics as an undergraduate at the California Institute of Technology. He did some graduate work in mathematics and computer science at the University of Wisconsin-Madison. He has worked in the computer industry for the past thirteen years and is currently with Microcomputer Solutions. Several years ago, while at home caring for his children, Mr. Gray watched an educational television show on mathematics. This show rekindled his interest in mathematics and ultimately led to his article about Cantor's work.

MILTON SOBEL is a product of CCNY in the Depression Era. His first job was with Dr. Ed Deming at the Census Bureau in 1940. He served with General George Patton (3rd Army) in World War II; he then completed his Ph.D. with Professor A. Wald in Mathematical Statistics at Columbia University (1951). After a six-year stint at Bell Telephone Laboratories, he went to the University of Minnesota and in the mid 1970's went further west to UCSB, where he allegedly retired in 1990. Actually he is still active and presently on recall, teaching at UCSB. His hobbies are chess, topical stamp collecting and a rare math book collection. Most of the above was made possible by his dear wife Florence, their 3 children, Marc, Judy and Eric, the 2 + grandchildren and his mother Tillie (age 100).

K. S. FRANKOWSKI was born in Poland. He finished music and regular high school, was one of the winners of Poland's first Mathematical Olympiad, studied music and mathematics. He completed a Master's degree in Mathematics at Warsaw University, and a Ph.D. in Applied Mathematics at the Hebrew University of Jerusalem under Professor Pekeris of the Weizmann Institute of Science. Frankowski is in the University of Minnesota's Computer Science Department. His primary interests are in applied physics and statistics, especially Dirichlet methods.

PAUL HALMOS has three degrees from the University of Illinois; soon after getting the last one he became, for a couple of years, assistant to John von Neumann. Since then he has taught at many universities (including Chicago, Michigan, and Indiana) and has visited many others (including Miami, Montevideo, Hawaii, Edinburgh, and Western Australia); he has been on the faculty of Santa Clara University since 1985. His mathematical interests include ergodic theory, algebraic logic, and operators on Hilbert space. Paul Halmos started learning, teaching, reading and writing (sixty years ago, at the University of Illinois), and he is still doing it (at Santa Clara University).

HUGH THURSTON was successively undergraduate at Trinity College, Cambridge; wartime cryptographer; research student (very abstract algebra) and teacher of mathematics (thirty-five years at the University of British Columbia); and is now a senior citizen.

Much teaching of calculus generated skepticism: hence the present article.

LOUIS GORDON completed a 1971 Ph.D. in Statistics at Stanford University under the direction of Bradley Efron. He has worked on the statistics of clinical trials while at ALZA Corporation, and on the statistics of oil and gas reserves while at the U.S. Department of Energy. He is now Professor of Mathematics at University of Southern California.

JACOB ELI GOODMAN received his Ph.D. at Columbia University in 1967, under the direction of Heisuke Hironaka. Since that year he has been at the City College, CUNY. In addition to algebraic geometry, he has worked in topological graph theory and discrete geometry, and has coauthored a series of papers with Richard Pollack focusing on the interplay between geometric and topological properties of configurations of points and arrangements of hyperplanes and their generalizations.

Professor Goodman was the recipient of a Fulbright grant during 1991–92. Along with Professor Pollack, he is coeditor in chief of *Discrete & Computational Geometry*, a journal of mathematics and computer science published by Springer-Verlag. When time permits, he improvises (classically) on the piano.

RICHARD POLLACK left C.C.N.Y. after a year devoted to bridge and chess. After a semester as a common laborer, he completed his B.S. at Brooklyn College, where he spent his spare time dancing in the Folk Dance Club. He received his Ph.D. at New York University in 1962 under the direction of Harold N. Shapiro, and has been on the faculty there ever since. After many years spent working in number theory, he began an exciting collaboration with Jacob E. Goodman in discrete geometry in 1978, and works in computational geometry as well. Professors Goodman and Pollack are currently working on an extension of the concept of convexity to sets of k -flats in d -dimensional affine space.

He is not the only mathematician in his family. His son, Daniel Pollack, and daughter-in-law, Tatiana Toro, are mathematicians too.

RAPHAEL WENGER received his Ph.D. in 1988 from the Department of Computer Science at McGill University under the supervision of David Avis. He held postdoctoral fellowships at the University of Montreal and the Center for Discrete Mathematics and Theoretical Computer Science and is currently an Assistant Professor in the Department of Computer and Information Sciences at Ohio State University. His main area of interest is computational and combinatorial geometry, and he has authored and co-authored numerous papers on the subject of geometric transversal theory.

TUDOR ZAMFIRESCU was born in Stockholm in 1944, went to the Kingdom of Romania at the age of 1, learned mathematics from his father, received his M.S. from the University of Bucharest in 1966, left Ceausescu's Romania, received his Ph.D. from the University of Bochum in 1968, received the Venia Legendi from the University of Dortmund in 1972, and became a professor there in 1977. He has held several visiting professorships in France, Italy, the United States, and the Netherlands.

Professor Zamfirescu has made essential contributions to the development of generic convexity. His main research areas are convex geometry and graph theory. He loved sports during his early youth, generalized sports during his later youth, and chaotic activism (raising children, organizing conferences, building buildings) in his maturity.

JEFFREY BOYLE earned his Ph.D. in Mathematics from the University of Iowa in 1984, spent three years as Visiting Professor at Michigan State University, and currently is Assistant Professor at the University of Wisconsin at La Crosse. He is a dedicated bicyclist (logging over 125,000 miles) and cross country skier. For the last seven years he has been trying to learn how to play the fiddle, but he is still not very good.

I have hardly ever known a mathematician who was capable of reasoning.

—Plato (ca 429–347 BC)
Republic, VII, 531.

Answer to Who are these mathematicians?
(p. 896)

Top, left to right: Pierre de Fermat (1601–1665), Christian Huygens (1629–1695), Jacob Bernoulli (1654–1705); Bottom, left to right: Joseph Louis Lagrange (1736–1813), Pierre Simon Marquis de Laplace (1749–1827).

They shared a common mathematical interest in the Sharing Problem (see the article in this issue by Sobel and Frankowski on page 833).

PROBLEMS AND SOLUTIONS

Edited by:

Richard T. Bumby, Fred Kochman and Douglas B. West

Proposed problems should be sent to the MONTHLY PROBLEMS address given on the inside front cover. Please include solutions and relevant references. Three copies of all items needed to evaluate the problem should be sent.

Solutions of published problems should arrive before April 30, 1995 at the MONTHLY PROBLEMS address given on the inside front cover. If possible, solutions should be typed with double spacing. Two copies suffice. Several solutions may be mailed together, but they should be on separate sheets of paper. The problem number and the solver's name and mailing address should appear on each solution. A mailing label should be included if an acknowledgment is desired.

The published solution is likely to be based on a solution that is complete and correct. Additional information, such as references to other appearances of the problem or its solution, is also welcome.

An asterisk () after the number of a problem, or part of a problem, indicates that no solution is currently available.*

PROBLEMS

10410. *Proposed by Frank Schmidt, Arlington, VA.*

Let G be a finite group. Define $a(G)$ to be the average order of an element of G . If $G \neq 1$, can $a(G)$ be an integer?

10411. *Proposed by Gord Sinnamon, University of Western Ontario, London, Ontario, Canada.*

Let R be the region inside the unit circle and above the line $x + y = 1$. Calculate

$$\iint_R \frac{1}{(\log x)^2 + (\log y)^2} \frac{dx dy}{xy}.$$

10412. *Proposed by Donald A. Darling, Newport Beach, CA.*

Find necessary and sufficient conditions on a non-increasing sequence a_1, a_2, \dots of positive real numbers so that, if b_1, b_2, \dots is a non-increasing sequence with $b_k \geq a_k$ for infinitely many k , then $\sum b_n = \infty$.

10413. *Proposed by Mirel Mocanu, University of Craiova, Craiova, Romania.*

Four disjoint (except for boundary points) equilateral triangles of sides a , b , c and d , are enclosed in a regular hexagon of unit side.

- (a) Prove that $3a + b + c + d \leq 4\sqrt{3}$.
- (b) When is $3a + b + c + d = 4\sqrt{3}$?
- (c)* Prove or disprove that $a + b + c + d \leq 2\sqrt{3}$.

10414. *Proposed by R. J. Simpson, Curtin University of Technology, Perth, Australia, and W. F. Smyth, McMaster University, Hamilton, Ontario, Canada.*

For a positive real number x , let

$$C(x) = \left\lceil \frac{x}{\lceil \sqrt{x} \rceil} \right\rceil + \lceil \sqrt{x} \rceil$$

and, for $x \geq 1$, let

$$F(x) = \left\lfloor \frac{x}{\lfloor \sqrt{x} \rfloor} \right\rfloor + \lfloor \sqrt{x} \rfloor.$$

- (a) Express $C(x)$ in a form that requires only one evaluation of a square root.
- (b) Express $F(x)$ in terms of $C(x)$.

10415. *Proposed by Edward Kitchen, Santa Monica, CA.*

Let \mathcal{A} be a triangle whose centroid is at the origin. Choose $k \in \mathbb{R}$, $k > 1$, and dilate one of the *Napoleon triangles* of \mathcal{A} by a factor of $-k$ and the other by a factor of $k/(1 - k)$. Prove that \mathcal{A} is (simultaneously) perspective with both dilated triangles.

10416. *Proposed by Kwang-Wu Chen (student), National Chung Cheng University, Chia-Yi, Taiwan, Republic of China.*

The Bernoulli numbers B_n ($n = 0, 1, 2, \dots$) are defined by

$$\frac{t}{e^t - 1} = \sum_{n=0}^{\infty} \frac{B_n t^n}{n!},$$

which converges for $|t| < 2\pi$. Also, for each nonnegative integer n , the Bernoulli polynomial $B_n(x)$ is defined by

$$B_n(x) = \sum_{k=0}^n \binom{n}{k} B_{n-k} x^k.$$

For integer $m \geq 1$ and arbitrary constants α and β , prove

$$\sum_{k=0}^m \binom{m}{k} B_k(\alpha) B_{m-k}(\beta) = -(m-1)B_m(\alpha + \beta) + m(\alpha + \beta - 1)B_{m-1}(\alpha + \beta).$$

NOTES

(10415) The outer Napoleon triangle of a triangle \mathcal{A} is formed by building equilateral triangles on the outside of each edge of \mathcal{A} and taking the centroids of these triangles as vertices. The so-called theorem of Napoleon states that this triangle is equilateral. The inner

Napoleon triangle is obtained in the same way from equilateral triangles on the edges of \mathcal{A} facing into \mathcal{A} . More information about Napoleon triangles can be found in John E. Wetzel, "Converses of Napoleon's theorem", this MONTHLY, 99 (1992), 339–351. Two triangles are in perspective if the vertices can be labelled so that the lines joining corresponding vertices are concurrent.

SOLUTIONS

Uniform Convergence and Continuity of Complex Power Series

6080 [1976, 205]. *Proposed by R. N. Hevener Jr., University of South Carolina.*

A theorem of Abel states that if $\sum_{n=0}^{\infty} a_n z^n$ converges on the closed interval A , then (i) convergence is uniform on A , whence (ii) it determines a continuous function on A . Is either part of this theorem true if A denotes a closed disk instead of an interval? If we impose the additional hypothesis, trivially satisfied in Abel's theorem, that the function be continuous on the boundary of A , is either part true?

Solution by M. J. Pelling, Royal Air Force of Oman, Seeb, Sultanate of Oman. The answers are negative, except that if f is continuous on the boundary ∂A it also will be on A .

Theorem 1. *If $f(z) = \sum_{n=0}^{\infty} a_n z^n$ converges on $A : |z| \leq 1$, then f need not be continuous on ∂A and, a fortiori, may not be continuous or uniformly convergent on A .*

Proof. Let

$$\begin{aligned} g(\theta) &= \exp\left(-\theta^{-1} \sin^2(\theta^{-1})\right) & 0 < \theta \leq \pi \\ &= g(-\theta) & -\pi \leq \theta < 0 \\ &= 0 & \theta = 0, \end{aligned}$$

and let a_n be the Fourier cosine coefficients of g . Then, by Dini's test, $\sum_{n=0}^{\infty} a_n \cos n\theta$ converges to $g(\theta)$ for all $\theta \in [-\pi, \pi]$. Writing $\psi(\theta, t) = g(\theta + t) - g(\theta - t)$, the conjugate series partial sum

$$-\sum_{m=1}^n a_m \sin m\theta = (2\pi)^{-1} \int_0^\pi \psi(\theta, t) \cot(t/2) (1 - \cos nt) dt + o(1)$$

by [1, 45–47]). This converges as $n \rightarrow \infty$ for all θ since $\psi(\theta, t)/t$ is $L(0, \delta)$ for all θ . In fact, by the Riemann-Lebesgue lemma,

$$h(\theta) = -\sum_{n=1}^{\infty} a_n \sin n\theta = (2\pi)^{-1} \int_0^\pi [g(\theta + t) - g(\theta - t)] \cot(t/2) dt.$$

Hence $f(z) = \sum_{n=0}^{\infty} a_n z^n$ converges on A , with $f(e^{i\theta}) = g(\theta) - ih(\theta)$ on ∂A . But f is not continuous on ∂A at $z = 1$, since $g(\theta)$ is discontinuous at $\theta = 0$.

Theorem 2. *If $f(z) = \sum_{n=0}^{\infty} a_n z^n$ converges on $A : |z| \leq 1$, then f is continuous on A if and only if it is continuous on ∂A . However, f may be continuous but not uniformly convergent on ∂A , so may be continuous yet not uniformly convergent on A .*

Proof. Trivially continuity on A implies continuity on ∂A . Conversely, if f is continuous on ∂A and $f(e^{i\theta}) = \sum_{n=0}^{\infty} a_n e^{ni\theta}$, then the latter series is the Fourier series of $f(e^{i\theta})$: see

theorem 100 on p.91 of [1]. If $P(r, \theta)$ denotes the Poisson kernel

$$\frac{(1-r^2)}{2\pi(1-2r\cos\theta+r^2)} = (2\pi)^{-1} \left(1 + 2 \sum_{n=1}^{\infty} r^n \cos n\theta \right) \quad 0 \leq r < 1,$$

then direct calculation shows

$$f(z) = \sum_{n=0}^{\infty} a_n z^n = \sum_{n=0}^{\infty} a_n r^n e^{ni\theta} = \int_0^{2\pi} P(r, \theta - t) f(e^{it}) dt$$

for $r = |z| < 1$. Interchange of the order of summation and integration is valid here by absolute convergence of the series expansion of $P(r, \theta)$ and boundedness of $f(e^{it})$. It follows by the Proposition on p.930 of [2] that if $f_r(\theta) = f(re^{i\theta})$ then the family $\{f_r : 0 \leq r < 1\}$ is equicontinuous at each θ . But by Abel's theorem $f(re^{i\theta})$ is continuous in r for $0 \leq r \leq 1$ and each θ and it follows that $f(z) = f(re^{i\theta})$ is 2-dimensionally continuous at each boundary point $z = e^{i\theta}$. Hence $f(z)$ is continuous on A .

Finally, theorem 1.17 on p. 301 of [3], constructs a power series $f(z) = \sum_{n=0}^{\infty} c_n z^n$ regular for $|z| < 1$, continuous for $|z| \leq 1$, and convergent on $|z| = 1$ but nonuniformly on every arc of $|z| = 1$.

In this context it may be of interest to note that theorem 1.14 on p. 300 of [3] constructs a power series $\sum_{n=0}^{\infty} c_n z^n$ regular in $|z| < 1$ which is the restriction to $|z| < 1$ of a continuous function $f(z)$ on $|z| \leq 1$, but for which $\sum c_n$ diverges. In this example, if $s(n, z) = \sum_{m=0}^n c_m z^m$, then there is a subsequence $s(n_k, z)$ of these partial sums of the power series such that $s(n_k, z) \rightarrow f(z)$ as $k \rightarrow \infty$ uniformly in $|z| \leq 1$.

REFERENCES

1. G. H. Hardy and W. W. Rogosinski, *Fourier Series*, Cambridge Tract 38, Cambridge University Press, 1968.
2. J. Král and W. F. Pfeffer, "Poisson integrals of Riemann integrable functions", this MONTHLY 98 (1991), 929-931.
3. A. Zygmund, *Trigonometric Series*, vol. I, second ed., Cambridge University Press, 1959.

There is no record of any other solution being received.

Comparing Sums of Numbers with Equal Products

6667 [1991,766]. Proposed by George Baloglou and Phil Tracy, State University of New York, College at Oswego.

If $a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n$ are positive numbers such that

$$a_1 a_2 \dots a_n = b_1 b_2 \dots b_n \quad \text{and} \\ \sum_{1 \leq i < j \leq n} |a_i - a_j| \leq \sum_{1 \leq i < j \leq n} |b_i - b_j|,$$

(i)* prove that

$$\sum_{i=1}^n a_i \leq (n-1) \sum_{i=1}^n b_i, \quad \text{and}$$

(ii) show that the factor $n-1$ cannot be replaced by a smaller one.

Solution of (i) by Eugene A. Herman, Grinnell College, Grinnell, IA. The following proof uses only the geometric-arithmetic mean inequality. By reordering subscripts, we may assume that

$$0 < a_1 \leq a_2 \leq \dots \leq a_n, \quad 0 < b_1 \leq b_2 \leq \dots \leq b_n. \quad (1)$$

Furthermore, the problem is unchanged if every a_i and b_i is multiplied by a positive constant k . If we choose $k = \left(\sum_{i=1}^n b_i \right)^{-1}$, the problem becomes the following: From assumptions (1), (2), (3), and (4), where

$$\prod_{i=1}^n a_i = \prod_{i=1}^n b_i \quad (2)$$

$$\sum_{i=1}^n (2i - n - 1)a_i \leq \sum_{i=1}^n (2i - n - 1)b_i \quad (3)$$

$$\sum_{i=1}^n b_i = 1 \quad (4)$$

prove

$$\sum_{i=1}^n a_i \leq n - 1. \quad (5)$$

Clearly, $n > 1$ was intended, since the result is false for $n = 1$. Also, the proof is routine when $n = 2$; so henceforth we assume $n \geq 3$ and prove (5) with strict inequality. Using the facts that $a_{n+1-i} - a_i \geq 0$ when $2i \leq n + 1$ and $(2i - n - 1)b_i \leq 0$ when $2i \leq n + 1$, we deduce from inequality (3) that

$$(n - 1)(a_n - a_1) \leq \sum_{i \geq (n+2)/2}^n (2i - n - 1)b_i. \quad (3')$$

The conclusion will follow if (3) is replaced by the weaker inequality (3'). From inequality (3') and equality (4), we have

$$\begin{aligned} a_n - a_1 &\leq 1 - \sum_{i=1}^n b_i + \sum_{i \geq (n+2)/2}^n \frac{2i - n - 1}{n - 1} b_i \\ &= 1 - \left(\sum_{i \leq (n+1)/2} b_i + \sum_{i \geq (n+2)/2}^{n-1} \frac{2(n-i)}{n-1} b_i \right). \end{aligned} \quad (6)$$

From (1) and equality (2), we have

$$\begin{aligned} a_1^{n-1} a_n &\leq \prod_{i=1}^n a_i \leq \prod_{i=1}^{n-1} b_i \quad (\text{since } b_n < 1) \\ &= \left(\prod_{i \leq (n+1)/2} b_i \right) \left(\prod_{i \geq (n+2)/2}^{n-1} \frac{2(n-i)}{n-1} b_i \right) \left(\prod_{i \geq (n+2)/2}^{n-1} \frac{n-1}{2(n-i)} \right) \\ &= \left(\prod_{i \leq (n+1)/2} b_i \right) \left(\prod_{i \geq (n+2)/2}^{n-1} \frac{2(n-i)}{n-1} b_i \right) \frac{(n-1)^k}{2^k k!} \end{aligned}$$

where $k = (n - 2)/2$ when n is even and $(n - 3)/2$ when n is odd. Therefore, by the geometric-arithmetic mean inequality,

$$\frac{a_1^{n-1} a_n}{(n-1)^k} \leq a_1^{n-1} a_n \frac{2^k k!}{(n-1)^k} \leq \frac{1}{(n-1)^{n-1}} \left(\sum_{i \leq (n+1)/2} b_i + \sum_{i \geq (n+2)/2}^{n-1} \frac{2(n-i)}{n-1} b_i \right)^{n-1}$$

Hence, by inequality (6),

$$a_1 \left(\frac{a_n}{(n-1)^k} \right)^{\frac{1}{n-1}} (n-1) \leq 1 - a_n + a_1, \text{ or}$$

$$a_1 \left[\left(\frac{a_n}{(n-1)^k} \right)^{\frac{1}{n-1}} (n-1) - 1 \right] \leq 1 - a_n. \quad (7)$$

If $a_n < (n-1)/n$, then $\sum_{i=1}^n a_i \leq na_n < n-1$. Thus we may assume that $a_n \geq (n-1)/n$, which implies

$$\left(\frac{a_n}{(n-1)^k} \right)^{\frac{1}{n-1}} (n-1) \geq \left(\frac{(n-1)^{n-k}}{n} \right)^{\frac{1}{n-1}} > 1.$$

So we see that in (7) the expression in square brackets is positive and hence, by inequality (7), that $1 - a_n > 0$. Since $\sum_{i=1}^n a_i \leq a_1 + (n-1)a_n$, the conclusion (5) holds with strict inequality if

$$\frac{1 - a_n}{\left(\frac{a_n}{(n-1)^k} \right)^{\frac{1}{n-1}} (n-1) - 1} + (n-1)a_n < n-1$$

(again using (7)) which we may rewrite successively as follows:

$$\begin{aligned} \frac{1}{n-1} &< \left(\frac{a_n}{(n-1)^k} \right)^{\frac{1}{n-1}} (n-1) - 1, \\ \left(\frac{n}{(n-1)^2} \right)^{n-1} &< \frac{a_n}{(n-1)^k}, \\ \frac{n^{n-1}}{(n-1)^{2n-2-k}} &< a_n. \end{aligned} \quad (8)$$

But the assumption $a_n > (n-1)/n$ implies

$$\frac{(n-1)^{2n-2-k} a_n}{n^{n-1}} \geq \frac{(n-1)^{2n-1-k}}{n^n}.$$

Thus to show that (8) holds, it suffices to prove that

$$n^n < (n-1)^{2n-1-k}.$$

This is easily confirmed when $n = 3$. When $n \geq 4$, $n < (n-1)^{\frac{3}{2}}$ and so

$$n^n < (n-1)^{\frac{3}{2}n} \leq (n-1)^{2n-1-k}.$$

This completes the proof of part (i).

Solution of (ii) by the editors. Take $a_1 = N^{-(n-1)}$ and $a_2 = a_3 = \dots = a_n = N$ for N large and positive; also $b_1 = b_3 = \dots = b_{n-1} = (N+1)^{-1/(n-1)}$ and $b_n = N+1$. Then

$$\sum_{1 \leq i < j \leq n} |a_i - a_j| < (n-1)N < \sum_{1 \leq i < j \leq n} |b_i - b_j|$$

so that the hypotheses of part (i) are satisfied; but also

$$\sum_{i=1}^n a_i > (n-1)N$$

and, for N sufficiently large,

$$\sum_{i=1}^n b_i < N+2,$$

so that

$$\sum_{i=1}^n a_i > \frac{(n-1)N}{N+2} \sum_{i=1}^n b_i.$$

Letting N grow large shows that the constant $n-1$ in part (i) cannot be improved.

Solved also by the proposers.

Reducing a matrix to zero by S E X changes

10216 [1992, 362]. *Proposed by G. Bennett, Indiana University, Bloomington, IN.*

Let $A = (a_{i,j})$ be an m by n matrix with integer entries. A set of locations, H , in A is called an “echelon” if, whenever $(k, l) \in H$, $i \leq k$ and $j \leq l$, one has $(i, j) \in H$. Consider the family of operations

$\mathbf{S}_{i,j}$: subtract 1 from $a_{i,j}$; add 1 to $a_{i+1,j}$

$\mathbf{E}_{i,j}$: subtract 1 from $a_{i,j}$; add 1 to $a_{i,j+1}$

$\mathbf{X}_{i,j}$: subtract 1 from $a_{i,j}$

(for all values of i and j for which the operations can be defined). Show that there is a sequence of these operations reducing A to the zero matrix if and only if

$$\sum \{a_{i,j} : (i, j) \in H\} \geq 0$$

for every echelon H .

Solution by Richard Stong, Rice University, Houston, TX. Let $f(H) = \sum_{(i,j) \in H} a_{i,j}$, and let $(*)$ be the condition that $f(H) \geq 0$ for every echelon H . Since no operation increases $f(H)$, $(*)$ is necessary. To see that $(*)$ is sufficient, note that each operation lowers $\sum_H f(H)$, where the sum is over all echelons $H \subseteq A$. Also, if $(*)$ holds, then the sum is a nonnegative integer and vanishes only if A is the zero matrix. Hence it suffices to show that some operation can be performed without destroying $(*)$ whenever A is nonzero and satisfies $(*)$.

Suppose $(*)$ holds. Call H a *zero echelon* if $f(H) = 0$. If H_1 and H_2 are zero echelons,

$$0 = f(H_1) + f(H_2) = f(H_1 \cap H_2) + f(H_1 \cup H_2).$$

Both terms on the right are nonnegative, hence both are zero. In other words, the set of zero echelons is closed under intersection and union.

Let $a_{i,j}$ be a positive entry of A . If (i, j) is not contained in any zero echelon, then we can apply $\mathbf{X}_{i,j}$. Otherwise, let H be the smallest zero echelon containing (i, j) , i.e., the intersection of all zero echelons containing (i, j) . One of the locations $(i+1, j)$ and $(i, j+1)$ must be in H , since otherwise $H' = H - \{(i, j)\}$ is an echelon with $f(H') = f(H) - a_{i,j} < 0$. By symmetry, we may assume $(i+1, j) \in H$. Now $\mathbf{S}_{i,j}$ can be performed to preserve $(*)$ unless there is a zero echelon \hat{H} that contains (i, j) but not $(i+1, j)$. In that case $H \cap \hat{H}$ would be a zero echelon containing (i, j) , contradicting the minimality of H .

Solved also by P. J. Anderson (Canada), D. M. Bloom, D. Callan, R. J. Chapman (U. K.), T. Hesterberg, S. Kanetkar, K. S. Kedlaya (student), O. P. Lossers (The Netherlands), M. Mócsy (Hungary), A. Nijenhuis, University of Wyoming Problem Circle, and the proposer.

The Determinant of an LCM Matrix

10232 [1992, 571]. *Proposed by Serge Zakharov, Tumen State University, Tumen, Russia.*

Let M_n be the n by n matrix whose (i, j) -entry is $\text{lcm}(i, j)$. Evaluate $\det(M_n)$.

Solution by Richard Holzsager, The American University, Washington, DC. One form for the answer is

$$\det(M_n) = n! \prod (1 - p)^{\lfloor n/p \rfloor},$$

where the product is taken over all primes p .

A function f defined on the positive integers is *multiplicative* if $f(mn) = f(m)f(n)$ whenever m and n are relatively prime. It is a standard result (found in most texts on

Elementary Number Theory) that if f is multiplicative, then so is the associated function g defined by $g(m) = \sum_{d|m} f(d)$. Let f be the multiplicative function defined by $f(p^n) = (1-p)/p^n$ for each prime power p^n , and let g be its associated multiplicative function. Since $g(p^n) = 1 + (1-p)/p + (1-p)/p^2 + \cdots + (1-p)/p^n = 1/p^n$, we conclude that $g(m) = 1/m$ for each positive integer m .

Now let N be the diagonal matrix $\text{diag}(1, 2, \dots, n)$, and let A be the n by n matrix whose (i, j) -entry is $1/\gcd(i, j)$. Then $M_n = NAN$. Next let B be the n by n matrix whose (i, j) -entry is $f(i)$ if $i|j$ and 0 otherwise, and let C be the n by n matrix whose (i, j) -entry is 1 if $j|i$ and 0 otherwise. Then the (i, j) -entry of the product CB is

$$\sum_{d=1}^n c_{id} b_{dj} = \sum_{d|i} b_{dj} = \sum_{d|\gcd(i, j)} f(d) = g(\gcd(i, j)) = 1/\gcd(i, j).$$

Thus $CB = A$.

We conclude that $\det(M_n) = \det(N)^2 \det(B) \det(C)$. Since the matrices N , B , and C are respectively diagonal, upper triangular, and lower triangular, we have $\det(N) = n!$, $\det(B) = \prod_{i=1}^n f(i)$, and $\det(C) = 1$. Thus $\det(M_n) = (n!)^2 \prod_{i=1}^n f(i)$.

Finally, define the function h by $h(m) = mf(m)$, so $\det(M_n) = n! \prod_{i=1}^n h(i)$. Now h is multiplicative, with $h(p^n) = 1-p$ for each prime power p^n . Thus $h(m) = \prod (1-p)$, where the product is taken over all primes dividing m . This yields the formula claimed, because for each prime p there are $\lfloor n/p \rfloor$ multiples of p in the interval $[1, n]$.

Editorial comment. This determinant appears frequently in the literature, including the textbook, I. Niven and H. S. Zuckerman, *An Introduction to the Theory of Numbers* (3rd ed.), Wiley, 1972. In this edition, Theorem 3 on p.266, gives the result in the form $\det(M_n) = \prod_{k=1}^n \phi(k) \prod_{p|k} (-p)$. Several readers traced the determinant to H. J. S. Smith, "On the value of a certain arithmetic determinant", *Proc. London Math. Soc.* 7 (1875-76), 208–212 (or *Collected Works* II, 161–165). It also appears in works that extend the theory of such matrices, such as S. J. Beslin, "Reciprocal GCD matrices and LCM matrices", *Fibonacci Quarterly* 29(1991), 271–274 and K. Bourke & S. Ligh, "On GCD and LCM matrices", *Linear Algebra and Appl.* 174(1992), 65–74.

Solved also by D. Alvis, D. W. Bailey, K. Bourque, D. Callan, S. R. Cavior, R. J. Chapman (U. K.), J. Duncan & M. Maliakas, P. Haukkanen (Finland), R. H. Jeurissen (The Netherlands), S. Kanetkar, N. Kang (student, Korea), S. P. Kishore & S. V. Singh (India), N. Komanda, J. H. Lindsey II, O. P. Lossers (The Netherlands), R. Martin (student), R. A. Mena, M. Mócsy (Hungary), A. Nijenhuis, J. Oaks, P. Schaefer, H. J. Seiffert (Germany), A. Yanushka, J. Zurek (France), USA Mathematical Olympiad Program, University of Wyoming Problem Circle, and the proposer.

A Stability Criterion

10271 [1992, 958]. Proposed by Victor I. Kostin, Institute of Mathematics, Novosibirsk, Russia.

Let A be a skew-hermitian N by N matrix with N distinct eigenvalues. Let b be a column vector with nonzero projections on each eigenvector of A . Prove that all eigenvalues of the $(N+1)$ by $(N+1)$ matrix

$$\begin{bmatrix} A & b \\ -b^* & -1 \end{bmatrix}$$

have negative real parts.

Solution by Dario Fasino, Università degli studi di Udine, Udine, Italy. A more general result holds.

Proposition. Let M be a complex matrix such that:

1. its hermitian part $H_1 = (M + M^*)/2$ is negative semidefinite;

2. if a vector v belongs to the kernel of H_1 , then v is not an eigenvector of $H_2 = (M - M^*)/2$, the skew-hermitian part of M .

Then, every eigenvalue of M has negative real part.

Proof. Let λ be any eigenvalue of M and v one of its eigenvectors: $Mv = (H_1 + H_2)v = \lambda v$. Since $H_1 = H_1^*$ and $H_2 = -H_2^*$, there exist real numbers c_1 and c_2 such that $v^* H_1 v = c_1 v^* v$ and $v^* H_2 v = ic_2 v^* v$. Then $\lambda = v^* M v / (v^* v) = c_1 + ic_2$ and $\Re(\lambda) = c_1 \leq 0$ by assumption 1. On the other hand, if $c_1 = 0$, then $H_1 v = 0$ and $Mv = H_2 v = \lambda v$, which is in contradiction with assumption 2. Thus $\Re(\lambda) < 0$.

We can apply this proposition to the given matrix $M = \begin{pmatrix} A & b \\ -b^* & -1 \end{pmatrix}$; indeed, in this case, $H_1 = \begin{pmatrix} A & 0 \\ 0^* & -1 \end{pmatrix}$, which is negative semidefinite. Moreover, let $v = (v_1, \dots, v_N, 0)^*$ be any nonzero vector in the kernel of H_1 and let $u = (v_1, \dots, v_N)^*$. Since $H_2 = \begin{pmatrix} A & b \\ -b^* & 0 \end{pmatrix}$, $H_2 v = \lambda v$ holds iff the relations $Au = \lambda u$ and $-b^* u = 0$ are both true, but $b^* u$ is not zero whenever u is an eigenvector of A . Hence v is not an eigenvector of H_2 . The hypothesis that the eigenvalues of A are distinct is not needed since, whenever A has an eigenspace of dimension greater than one, the required condition on the projections of b cannot hold.

Editorial comment. An obvious misprint in the original statement has been corrected in the above statement of the problem. A generalization, as in the selected solution, was also given by Michael K. Kinyon, C. R. Rosentrater, and the University of Wyoming Problem Circle.

Solved also by R. J. Chapman (U. K.), D. Jespersen, M. K. Kinyon, M. Marcus, C. R. Rosentrater, F. Schmidt, GCHQ Problem Solving Group (U. K.), University of Wyoming Problem Circle, and the proposer. One incomplete solution was received.

Convex Functions on Convex Sets

10283 [1993, 184]. *Proposed by Feng Luo and Richard Stong, University of California, Los Angeles, CA.*

Let \mathbf{D} be a convex polygonal region in the plane and let f be a bounded convex (and hence continuous) function on the interior of \mathbf{D} .

(a) Show that f extends to a continuous function on all of \mathbf{D} .

(b) Show that the analogous result does not hold if \mathbf{D} is the unit disk.

Solution by John Rainwater, University of Washington, Seattle, WA. This problem was answered in David Gale, Victor Klee and R. T. Rockafellar, "Convex functions on convex polytopes", *Proc. Amer. Math. Soc.* 19 (1968), 867–873, where it was shown (among more general results) that a bounded convex function on the relative interior of an n -dimensional polytope \mathbf{D} has a unique extension to a continuous convex function on \mathbf{D} . Moreover, this property characterizes polytopes among bounded convex sets. This answers both parts of the problem. A concrete example for the unit disk $\mathbf{D}_1 = \{(x, y) : x^2 + y^2 \leq 1\}$ is the function

$$f(x, y) = \frac{y^2}{1 - |x|} \quad (|x| < 1).$$

This is convex since it is the maximum of $y^2/(1 - x)$ and $y^2/(1 + x)$, each of which has a positive semidefinite Hessian, hence is convex. However, $f(x, y) \leq 2$ on \mathbf{D}_1 with $f(x, y) \rightarrow 2$ as $(x, y) \rightarrow (1, 0)$ along the circle, while $f(x, 0) = 0$ for $|x| < 1$.

Editorial comment. Dale Varberg and Wayne Roberts gave a related example, and remarked that composition with a rotation would lead to functions $f_\theta(x, y)$ with discontinuities at $\pm(\cos \theta, \sin \theta)$. Functions of the form

$$\sum 2^{-n} f_{\theta_n}(x, y)$$

could then be formed with a dense set of points of discontinuity on the boundary of D_1 .

Solved also by M. V. Bjelica (Yugoslavia), Y. Diao, H. von Eitzen (Germany), H. Hanche-Olsen, J. Kane, I. Kastanas, A. D. Melas (Greece), D. Varberg & W. Roberts, GCHQ Problem Solving Group (U. K.), and the proposers.

Collaborating editors: *David F. Appleyard, Paul T. Bateman, Bruce C. Berndt, Duane M. Broline, Barry W. Brunson, Frank S. Cater, Gulbank D. Chakerian, Underwood Dudley, Gerald A. Edgar, Michael A. Filaseta, Ira M. Gessel, Richard A. Gibbs, Jerrold R. Griggs, Douglas A. Hensley, John R. Isbell, Mourad E. H. Ismail, Murray Klamkin, Daniel J. Kleitman, Frederick W. Luttman, Frank B. Miles, Richard Pfeifer, Stephen L. Portnoy, J. O. Shallit, John Henry Steelman, Kenneth B. Stolarsky, David E. Tepper, Douglas B. Tyler, Daniel Ullman, and William E. Watkins.*

For all their wealth of content, for all the sum of history and social institution invested in them, music, mathematics, and chess are resplendently useless (applied mathematics is a higher plumbing, a kind of music for the police band). They are metaphysically trivial, irresponsible. They refuse to relate outward, to take reality for arbiter. This is the source of their witchery.

—G. Steiner

REVIEWS

Edited by **Darrell Haile**

Indiana University, Bloomington IN 47405

Fleeting Footsteps: Tracing the Conception of Arithmetic and Algebra in Ancient China. By Lam Lay Yong and Ang Tian Se. Singapore (World Scientific Publishing Co., Inc.) 1992. XVI + 199 pp. U.S. \$24, £ 15.

Reviewed by **Frank Swetz**

The fascinating title of this book owes its origin to a remark/query posed by the nineteenth century mathematical historian Florian Cajori concerning the “Hindu-Arabic” numerals. Cajori was puzzled as to why scholars could not agree on the course of these numerals, “their fleeting footsteps as they migrated from land to land.” (Cajori, 1896). Lam Lay Yong, Professor of Mathematics at the National University of Singapore and Ang Tian Se, formerly Professor of Chinese Studies at the University of Malaya and now teaching at Edith Cowan University, Australia, attempt to trace the elusive origin of these “footsteps” back to ancient China. The authors offer and embellish the premise that the concept of the “Hindu-Arabic” numerals can be found in the rod numerals used by the early Chinese. In building their case, Lam and Ang discuss the structure, (base 10, positional, etc.), and form of the rod numerals as well as describe the computational rod algorithms used to perform the four basic operations and extract numerical roots of numbers. Physical configurations of small rods laid out on a counting board served as a computing device while the written record of the rod configurations served as numerals. This system was highly efficient and eventually allowed for an operational designation of negative numbers, black rods for negative numbers and red rods for positive, and an extension to work with decimal fractions. Early Chinese computers also developed computational algorithms to perform what today would be recognized as algebraic procedures: solving a proportion by use of the “rule of three”, using “false position” to obtain a solution to a simple linear equation and solving a system of equations by using a tabular or matrix solution scheme. This discussion of rod arithmetic and algebraic procedures is primarily based on the contents of the fifth century mathematics text *Sun Zi suanjing* [The Mathematical Classic of Sun Zi]. A complete translation of this work is provided. Thus in the *Fleeting Footsteps* the authors have set themselves an ambitious and multifaceted agenda, namely to: provide a translation and presentation of the contents of *Sun Zi suanjing*; use the contents of this ancient classic to furnish an explanation of rod numeral arithmetic and algebra and then, on the basis of having established the mathematical superiority of rod numerals and computation, contend the case for a Chinese origin of the “Hindu-Arabic” numeral system.

This book presents the first, complete, English language translation of *Sun Zi suanjing*, one of the *Ten Mathematical Manuals* of ancient China (pp. 151–182). The dating of Master Sun’s classic is based on internal evidence and has been

approximated to the years between 280 AD and 473 AD. Sun Zi's specific identity also remains clouded and he is believed to have been a minor state official or a Buddhist monk. His brief mathematical text consists of a preface and three chapters in which he provides instructions for the use of computing rods, tables of weights and measures relevant to the time, and a collection of sixty-four mathematical problems and their solution schemes. The problems reflect the needs of daily life in fifth-century China. The text's intended audience was scholars and officials and not the "common people" as suggested by the authors (pp. 127). Although, in general, this early Chinese work remains unknown in the West, one of its problems has become famous and is widely recognized as the origin of the "Chinese Remainder Theorem" (Swetz, 1979). Now, this whole collection of intriguing problems which, in many ways, is reminiscent of Alcuin of York's *Propositiones ad acuendos jeevenes* (ca. 732), is available to a wide audience. This translation has supplied mathematical historians and others who are non-readers of Chinese with a true treasure! For example, problem thirty-four of chapter two tests the numerical endurance of a computer by requiring "Now there is sighted 9 embankments outside; each embankment has 9 trees; each tree has 9 branches; each branch has 9 nests; each nest has 9 birds; each bird has 9 young birds; each young bird has 9 feathers; each feather has 9 colors. Find the quantity of each." It is found there are 43,046,721 colors (pp. 181–2). But there are also other problems more representative of a particular Chinese mystical nature—"Now there is a pregnant woman whose age is 29. If the gestation period is 9 months, determine the sex of the unborn child". Through a number manipulation based on Yin-Yang principles, the child is found to be a male (pp. 182). In their analysis of rod procedures and operations, Lam and Ang rely extensively on the use of these problems. An appendix is provided associating each particular problem with its textual discussion within the book, a very valuable reference feature (pp. 183–84). Chapter eight of *Fleeting Footsteps* is devoted to a discussion of the economic, sociological and political aspects of the problem collection. These insights into Chinese life consider factors such as the use of barter, taxation rates and evidence of existing social unrest. This examination is an excellent feature of the book and I regret it was not longer and more detailed in its considerations. In brief, the presentation and discussion of the contents of *Sun Zi suanjing* is well done.

The existence and use of computing rods in China can be traced back to the Warring States period of history (475–221 BC). These rods were made of wood, bone or even ivory, depending on the rank and privileges of their user. They were approximately 0.2 cm thick and 8 cm long. In the hands of a skilled calculator, they were a powerful computing tool, in many respects equivalent to the hand calculator of today. The use of such rods dominated the Chinese computing scene up until the 17th century when they were replaced by the more egalitarian abacus. Discussions on rod algorithms and computing procedures occupy chapters three through five and present lucid explanations of the workings of rod arithmetic and simple algebra. This material provides the most extensive and detailed account of ancient Chinese rod mathematics available in the English language and opens another window of understanding for the linguistically limited reader.

Thus after recognizing some "footsteps" in the conceptualization of the "Hindu-Arabic" numeral system, the senior author, Lam Lay Yong, proposes on the basis of the evidence examined, that the path of these "footsteps" originated in China—"the Hindu-Arabic numeral system had its origins in the Chinese rod numeral system" (pp. 134). Of course this claim is not new. Alexander Wylie, one of the first modern, western observers of the Chinese mathematical scene noted

the possibility of a Chinese origin for the Hindu-Arabic numeral system (Wylie, 1897). The conjecture was echoed by early 20th century scholars (Fellows, 1921). In most recent times, this claim has been made most succinctly in the works of Wang Ling and Joseph Needham. In a 1958 paper delivered in Adelaide, Wang presented a detailed case for a Sino origin of the “Hindu-Arabic” numerals and pointed to the strong possibility of westward transmission to India. Wang’s theory was further amplified in his collaborative work with Joseph Needham. *Science and Civilization in China*, vol. 3, devotes several pages (pp. 146–150) to this very issue and the phenomena of “stimulus diffusion”. Needham’s work clearly indicates the need for further research clarification as to the status of early Hindu mathematics and the possibility of cultural transmissions. It is exactly this research that must be undertaken to strengthen the claim for a Chinese genesis of our numeral system and, unfortunately, it is exactly this research that is lacking in *Fleeting Footsteps*. What was the status of ancient Indian mathematics during the Warring States period of Chinese history? How were the numerals used in ancient India? Could the Chinese have obtained their mathematical knowledge from India?—after all, Buddhism was an intellectual import from China’s western neighbor. These are some of the issues and questions that must be addressed in positing a claim of a Chinese origin for the “Hindu-Arabic” numeral system and they remain missing footsteps in the path this book has taken.

Despite the inability to develop and strengthen its major premise, *Fleeting Footsteps* is a valuable resource for understanding early Chinese mathematics. Its sections devoted to the *Sun Zi suanjing* and its problems and to rod calculations provide historical information and insights previously unknown to a large audience. Although the “Fleeting Footsteps” themselves remain elusive, the book that bears this name is highly recommended for general reading and library acquisition.

REFERENCES

- Cajori, Florian, *A History of Elementary Mathematics*, New York: Macmillan Co., (1896).
 Eastlake, F. W., Finger Reckoning in China, *China Review* 9 (1880) 250–251.
 Fellows, Albion, Chinese mathematics—the Oldest, *Chinese Review* 1 (1921) 214–215.
 Needham, Joseph, *Science and Civilization in China* vol. 3, Cambridge: Cambridge University Press, (1959).
 Swetz, Frank, The evolution of mathematics in ancient China, *The Mathematics Magazine*, 52 (1987) 10–19.
 Wang Ling, The Chinese origin of the decimal place—value system in the notation of numbers and the possibility of its transmission to India, Paper presented at International Congress, Adelaide, Australia, (20–27 August, 1958) 5.
 Wylie, Alexander, *Chinese Researches*, Shanghai: Mission Press, (1897).

Department of Mathematics
Pennsylvania State University
Middletown, PA 17057-4898

Answer to Picture Puzzle (p. 857)

Philip Harman and Aurel Wintner

TELEGRAPHIC REVIEWS

Edited by **Arnold Ostebee and Paul Zorn**

with the assistance of the Mathematics Departments of
Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

T : Textbook	P : Professional Reading	1-4 : Semester
C : Computer Software	L : Undergraduate Library	** : Special Emphasis
S : Supplementary Reading	13 : Grade Level	?? : Questionable

Readers are advised that price information is subject to change. Selected books and software packages receive a second, more extensive review in the *Monthly*.

Books and software submitted for review should be sent to *Book Reviews Editor, American Mathematical Monthly, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098*.

General, P. *The Gelfand Mathematical Seminars, 1990–1992*. Eds: L. Corwin, I. Gelfand, J. Lepowsky. Birkhäuser, 1993, x + 235 pp, \$59.50. [ISBN 0-8176-3689-7] 15 papers on miscellaneous topics.

Reference, P, C. *Russian-English, English-Russian Dictionary on Probability, Statistics, and Combinatorics*. K.A. Borovkov. SIAM, 1994, viii + 154 pp, \$47.50 (P), with disk. [ISBN 0-89871-316-1]

Reference, P. *McGraw-Hill Dictionary of Scientific and Technical Terms, Fifth Edition*. Sybil P. Parker. McGraw-Hill, 1994, xvii + 2242 pp, \$110.50. [ISBN 0-07-042333-4]

Mathematics Appreciation, T(13: 1, 2). *Patterns in Mathematics: Problem Solving from Counting to Chaos*. Jack R. McCown, Michael A. Sequeira. PWS, 1994, xx + 874 pp. [ISBN 0-534-18786-2] "Math for liberal arts:" problem solving, applications, history, exploration problems, motivational quotes. Covers sets, counting, probability, statistics, Markov chains, fractals, and logic. As in Sherlock Holmes novels, sections start with an "opening case." HD

Mathematics Appreciation, S, L*.** *Build Your Own Polyhedra*. Peter Hilton, Jean Pedersen. Addison-Wesley, 1994, 175 pp, (P). [ISBN 0-201-49096-X] A mix of origami and mathematics. Thorough, detailed instructions for building many polyhedra. Illustrations and relevant geometry at high school level. Great for math ed courses, summer institutes. (1988 edition, TR, January 1989.) DP

Mathematics Appreciation, S.** *Classic Math: History Topics for the Classroom*. Art

Johnson. Dale Seymour Pub, 1994, vii + 172 pp, (P). [ISBN 0-86651-690-5] Wonderful supplement for various courses. Five chapters: Quote of the Week, Event of the Week, Historical Problems, History of Mathematical Symbols, History of Mathematical Terms. Each consists of several short sections aiming to stimulate discussion, convey human side of mathematics. Intended for grades 7–12. DP

Recreational Mathematics, S(13). *Speed Mathematics Simplified*. Edward Stoddard. Dover, 1994, x + 271 pp, \$6.95 (P). [ISBN 0-486-27887-5] Unaltered reprint of 1965 edition. Explains tricks, many based on abacus principle, for rapid mental arithmetic. HD

Elementary, P, L*. *Algebra*. I.M. Gelfand, A. Shen. Birkhäuser, 1993, 153 pp, \$24.50. [ISBN 0-8176-3737-0] Elementary algebra text for high school students in Gelfand's Mathematical School by Correspondence. Spare exposition; challenging but intriguing problems—many with solutions. A pedagogical gem. AO

Education, P. *The Legacy of Hans Freudenthal*. Ed: Leen Streefland. Kluwer Academic, 1993, 164 pp, \$60. [ISBN 0-7923-2653-9] Reprinted from *Educational Studies in Mathematics* 25 (1–2) 1993.

Education, P. *Didactics of Mathematics as a Scientific Discipline*. Eds: Rolf Biehler, et al. Math. Educ. Lib., V. 13. Kluwer Academic, 1994, ix + 467 pp, \$150. [ISBN 0-7923-2613-X] Papers on preparing mathematics for students, teacher education and research on teaching, interaction in the classroom, technology and mathematics education, psychol-

ogy of mathematical thinking, differential didactics, history and epistemology of mathematics and mathematics education, cultural framing of teaching and learning mathematics.

Education, S(15–17). *Algebra Experiments I: Exploring Linear Functions.* Mary Jean Winter, Ronald J. Carlson. Addison-Wesley, 1993, 105 pp, (P). [ISBN 0-201-81524-9] 17 experiments, each with teaching notes and lab pages. Stresses dependent vs. independent variables. Useful for pre- and in-service teachers. MW

Education, S(13–16). *Algebra Experiments II: Exploring Nonlinear Functions.* Ronald J. Carlson, Mary Jean Winter. Addison-Wesley, 1993, 108 pp, (P). [ISBN 0-201-81525-7] 14 experiments explore rational, exponential, and logarithmic functions. Includes programs for graphing calculators, Apple Basic. MW

Education, S(15–17). *Cooperative Informal Geometry.* Wade H. Sherard III, Dale Seymour Pub, 1995, 92 pp, (P). [ISBN 0-86651-799-5] 20 investigations, at middle school level, stress cooperative work groups, hands-on activities. With student worksheets, teacher notes. Useful as models for secondary teachers. MW

Logic, T(13–14: 1), L. *Logic, Sets, and Recursion.* Robert L. Causey. Ser. in Comp. Sci. Jones & Bartlett, 1994, x + 405 pp, \$49.95. [ISBN 0-86720-463-X] Careful, lucid introduction to propositional calculus, sets, relations, functions, predicate calculus. Stresses proofs and understanding; aims to fill gap in background of CS (and math) students. RM

Logic, P. *The Structure of Relation Algebras Generated by Relativizations.* Steven R. Givant. Contemp. Math., V. 156. AMS, 1994, xv + 134 pp, \$34 (P). [ISBN 0-8218-5177-2]

Foundations, T(15–16: 1), L. *Notes on Set Theory.* Yiannis N. Moschovakis. Undergrad. Texts in Math. Springer-Verlag, 1994, xiv + 272 pp, \$39. [ISBN 0-387-94180-0] A modern introduction. Covers interesting set-theoretic questions as well as role of set theory as foundation for mathematics: faithful representation of mathematical objects by structured sets, domains and the fixed point theorem, Aczel's antifounded universe. RM

Graph Theory, T(13–14: 1), S, L. *Introduction to Graph Theory.* Richard J. Trudeau. Dover, 1993, x + 209 pp, \$7.95 (P). [ISBN 0-486-67870-9] Nice, simple, enjoyable introduction to graph theory; an introduction to abstract mathematics for the "mathematically traumatized" (author's comment). Reprint of 1976 version published as *Dots and Lines* (TR's, March 1977 and February 1979). RM

Combinatorics, L. *Concrete Mathematics, Second Edition.* Ronald L. Graham, Donald E. Knuth, Oren Patashnik. Addison-Wesley, 1994, xiii + 657 pp, \$51.75. [ISBN 0-201-55802-5] Major change is new section on Zeilberger's extensions to Gosper's algorithm for evaluating indefinite sums. (First Edition, TR, May 1990; Extended Review, October 1991.) AO

Combinatorics, P. *Investigations in Algebraic Theory of Combinatorial Objects.* Eds: I.A. Farad'zev, et al. Math. & Its Applic., V. 84. Kluwer Academic, 1994, xi + 510 pp, \$269. [ISBN 0-7923-1927-3] Survey and research papers on cellular rings, distance-transitive graphs, amalgams and diagram geometries.

Discrete Mathematics, T(13–14: 1), L. *A Logical Approach to Discrete Math.* David Gries, Fred B. Schneider. Texts & Mono. in Comp. Sci. Springer-Verlag, 1993, xvi + 497 pp, \$44.95. [ISBN 0-387-94115-0]; *Instructor's Manual*, iv + 311 pp, (P). Unconventional, relatively sophisticated. Includes 10 chapters on logic and its applications to computer science. Also covers set theory, induction, relations and functions, number theory, combinatorics, recurrence relations, graphs. KES

Linear Algebra, T(14–15). *Linear Algebra: Gateway to Mathematics.* Robert Messer. HarperCollins, 1993, x + 403 pp, \$44. [ISBN 0-06-501728-5] Fairly conventional in outline and content; book's novelty lies mainly in various user-friendly features: informal remarks blocked off from the main text, two student projects for each chapter, numerous exercises, review exercises, chapter summaries. JS

Linear Algebra, T(13–14: 1). *Elementary Linear Algebra, Fifth Edition.* Stanley I. Grossman. Saunders College, 1994, xx + 758 pp, \$58.75. [ISBN 0-03-097354-6] Major changes include "calculator boxes"—instructions for matrix features on TI-85 and Casio fx-7700 GB; MATLAB tutorials and problems; a section on LU factorization. 130 pages longer than *Fourth Edition* (TR, October 1991), due mainly to MATLAB material. LC

Group Theory, P. *Geometric Group Theory, Volume 2: Asymptotic Invariants of Infinite Groups, M. Gromov.* Eds: Graham A. Niblo, Martin A. Roller. London Math. Soc. Lect. Note Ser., V. 182. Cambridge Univ Pr, 1993, vii + 295 pp, \$34.95 (P). [ISBN 0-521-44680-5] An extended essay by Mikhael Gromov, partly based on a lecture at a 1991 symposium at Sussex University.

Group Theory, T(17: 1). *Fast Fourier Transforms.* Michael Clausen, Ulrich Baum. Bib-

liographisches Institut & FA Brockhaus, 1993, 181 pp, DM 68. [ISBN 3-411-16361-5] Design and analysis of FFTs on finite groups; self-contained introduction to representation theory. Applications to graph theory, signal processing, statistics, data compression. TH

Algebra, S(18), P. *Fundamental Structures of Algebra and Discrete Mathematics*. Stephan Foldes. Wiley, 1994, xv + 344 pp, \$64.95. [ISBN 0-471-57180-6] Treats various algebraic structures, graphs, matroids, topological spaces, universal algebra, categories. Useful mainly as a concise summary of major results—exercises, examples, and motivation are too limited for beginners. JS

Algebra, P. *Three Papers on Algebras and Their Representations*. V.N. Gerasimov, N.G. Nesterenko, A.I. Valitskas. Transl. Ser. 2, V. 156. AMS, 1993, x + 195 pp, \$98. [ISBN 0-8218-7503-5]

Algebra, P. *Universal Algebra, Algebraic Logic, and Databases*. B. Plotkin. Math. and Its Applic., V. 272. Kluwer Academic, 1994, xv + 438 pp, \$189. [ISBN 0-7923-2665-2] Applies universal algebra and category theory to computer science, in this case to algebraic aspects of database theory. Employs many-sorted algebras and topoi as models; applies algebraic logic to the structure of databases (including those with fuzzy information). RM

Algebra, P. *Unconventional Lie Algebras*. Ed: Dmitry Fuchs. Adv. in Soviet Math., V. 17. AMS, 1993, x + 216 pp, \$105. [ISBN 0-8218-4121-1] 8 papers on representations and cohomology of Lie algebras that are infinite-dimensional, defined over fields of finite characteristic, Lie superalgebras, or quantum groups.

Algebra, P. *Representations of Algebras*. Eds: Vlastimil Dlab, Helmut Lenzing. Canadian Math. Soc. Conf. Proc., V. 14. AMS, 1993, xxv + 478 pp, \$81 (P). [ISBN 0-8218-6019-4] Proceedings of a 1992 conference at Carleton University.

Calculus, S(13). *Calculus: An Active Approach with Projects*. Stephen Hilbert, et al. Wiley, 1994, xi + 257 pp, \$20.95 (P). [ISBN 0-471-00316-6] Activities and projects to promote students' own development of calculus concepts. "Activities" are for in-class or homework use; 2–3 week team projects emphasize written presentation of results. Accompanies any calculus text; contains tear-out assignment pages. *Instructor's Guide* available. JNC

Calculus, T(13: 1–3). *Calculus with Analytic Geometry, Alternate Fifth Edition*. Roland E. Larson, Robert P. Hostetler, Bruce H. Edwards. DC Heath, 1994, xxii + 1243 pp [ISBN 0-669-

35336-1]. *Calculus with Analytic Geometry, Fifth Edition*. 1994, xxxviii + 1256 pp. [ISBN 0-669-35335-3] "Every portion . . . examined and revised in the spirit of [calculus] reform." More use of technology, more emphasis on communication skills, new applications. *Alternate Fifth Edition* has "late trigonometry," different treatment of limits, integral applications, exponential and log functions, vectors. (*Third Edition*, TR, May 1986.) AO

Calculus, T(13: 1–3), L. *Calculus of a Single Variable*. Thomas P. Dick, Charles M. Patton. PWS, 1994, xxiv + 831 pp. [ISBN 0-534-93936-8] One of the "new" texts. Contents not too unstandard; informal approach. Reorderings include early exponentials, parametric curves as application of derivatives, differential equations immediately after integration. Only antidifferentiation techniques are substitution and parts (partial fractions, trigonometric techniques, etc., in appendices). Contains such pleasant (and popular) morsels as splines and Bezier curves, Cantor set, iterative processes, a fixed-point theorem. Is "technology-aware," but assumes no specific technology. KS

Calculus, T(13: 1–3). *Calculus, Second Edition*. Richard A. Hunt. HarperCollins, 1994, xviii + 1165 pp, \$53.50. [ISBN 0-06-043046-X] Over 1700 new exercises: applications; conceptual and theoretical problems; graphing calculator exercises; writing exercises. AO

Real Analysis, T(15). *Elementary Classical Analysis, Second Edition*. Jerrold E. Marsden, Michael J. Hoffman. WH Freeman, 1993, xiv + 738 pp, \$47.95. [ISBN 0-7167-2105-8] Attractive option for a first course. Retains many attractive features of *First Edition* (TR, April 1975)—intuitive presentation of key theorems, followed at chapter end by formal proofs; close ties to applications. Length may be excessive for some—the derivative first appears on page 327. AWR

Differential Equations, S. *Differential Equations With Derive*. David C. Arney. MathWare (604 E. Mumford Dr., Urbana, IL 61801), 1993, x + 284 pp, \$23 (P). [ISBN 0-9623629-3-X] Examples and exercises that use Derive to solve or analyze ODE's. SK

Partial Differential Equations, P. *Computational Methods for Boundary and Interior Layers in Several Dimensions*. Ed: J.J.H. Miller. Boole Pr, 1991, vii + 225 pp. (P). [ISBN 0-906783-93-3; 0-906783-92-5] 10 papers survey developments in theory and application of numerical solutions of multi-dimensional problems with boundary and interior layers.

Partial Differential Equations, P. *Domain*

Decomposition Methods in Science and Engineering. Eds: Alfio Quarteroni, *et al.* Contemp. Math., V. 157. AMS, 1994, xxii + 484 pp, \$75 (P). [ISBN 0-8218-5158-6] Proceedings of a 1992 conference in Como, Italy.

Dynamical Systems, P. *Topics in Ergodic Theory.* Ya. G. Sinai. Math. Ser., V. 44. Princeton Univ Pr, 1994, vii + 218 pp, \$39.50. [ISBN 0-691-03277-7]

Numerical Analysis, T(15-16: 1, 2). *Numerical Analysis.* Vithal A. Patel. Saunders College, 1994, xvi + 652 pp, \$44. [ISBN 0-03-098330-4] Methods and algorithms for scientific computing: root finding, interpolation, differentiation, integration, systems of linear and nonlinear equations, linear algebra, ODEs, PDEs. Pseudocode algorithms; paper-and-pencil, computer exercises for each section. AO

Numerical Analysis, P. *Parallelism in the Numerical Integration of Initial Value Problems.* B.P. Sommeijer. CWI Tracts, V. 99. Centrum voor Wiskunde en Informatica, 1993, v + 195 pp, Dfl. 50 (P). [ISBN 90-6196-431-8] An introduction and 6 (republished) technical papers on parallel numerical methods for ODE's (both nonstiff and stiff).

Numerical Analysis, P. *Applications of Advanced Computational Methods for Boundary and Interior Layers.* Ed: J.J.H. Miller. Boole Pr, 1993, vii + 215 pp, (P). [ISBN 1-85748-002-3; 1-85748-001-5] 9 papers illustrate new computational techniques for numerical solution of singularly perturbed problems.

Operator Theory, P. *Harmonic Analysis and Boundary Value Problems in the Complex Domain.* Mkhitar M. Djrbashian. Oper. Theory: Adv. & Applic., V. 65. Birkhäuser, 1993, xiii + 256 pp, \$100. [ISBN 0-8176-2855-X]

Functional Analysis, P. *Multimedians in Metric and Normed Spaces.* E.R. Verheul. CWI Tracts, V. 91. Centrum voor Wiskunde en Informatica, 1993, x + 136 pp, Dfl. 40 (P). [ISBN 90-6196-420-2]

Functional Analysis, P. *Semigroups of Linear and Nonlinear Operations and Applications.* Eds: Gisèle Ruiz Goldstein, Jerome A. Goldstein. Kluwer Academic, 1993, 283 pp, \$119. [ISBN 0-7923-2560-5] Proceedings of a 1992 Curaçao conference/workshop.

Analysis, P. *A Lie Algebraic Study of Some Integrable Systems Associated with Root Systems.* J.K. Scholma. CWI Tracts, V. 96. Centrum voor Wiskunde en Informatica, 1993, 92 pp, Dfl. 30 (P). [ISBN 90-6196-426-1]

Analysis, P. *Analysis of and on Uniformly Rectifiable Sets.* Guy David, Stephen Semmes.

Math. Surveys & Mono., V. 38. AMS, 1993, xii + 356 pp, \$111. [ISBN 0-8218-1537-7]

Analysis, P. *Residue Currents and Bezout Identities.* Carlos A. Berenstein, *et al.* Progress in Math., V. 114. Birkhäuser, 1993, xi + 158 pp, \$49.50. [ISBN 0-8176-2945-9]

Algebraic Geometry, P. *Recent Advances in Real Algebraic Geometry and Quadratic Forms.* Eds: William B. Jacob, Tsit-Yuen Lam, Robert O. Robson. Contemp. Math., V. 155. AMS, 1994, viii + 404 pp, \$57 (P). [ISBN 0-8218-5154-3] Papers from a 1990-91 program at the University of California, Berkeley, and from a 1991 special session at the AMS winter meeting in San Francisco.

Differential Geometry, P. *Lie-Cartan-Ehresmann Theory.* Robert Hermann. Interdisc. Math., V. 28. Math Sci Pr, 1993, 283 pp, \$95. [ISBN 0-915692-44-9] Three approaches to differential geometry laid out in the author's unique style. Treats applications of geometric structures such as frame bundles, Lie structures, and variational calculus. Foreword contains author's opinions on current funding and state of mathematics in the U.S. DS

Differential Geometry, P. *Functions on Manifolds: Algebraic and Topological Aspects.* V.V. Sharko. Transl. of Math. Mono., V. 131. AMS, 1993, x + 193 pp, \$98. [ISBN 0-8218-4578-0]

Geometry, T*(15: 1), L. *Topics in Geometry.* Robert Bix. Academic Pr, 1994, x + 538 pp, \$59.95. [ISBN 0-12-102740-6] For prospective secondary teachers and other mathematics majors. Covers topics in, and emphasizes connections among, advanced Euclidean geometry, transformation geometry (including wallpaper groups), projective geometry, conic sections, hyperbolic and absolute geometry. Many exercises of varying difficulty. JNC

Geometry, S, P, L. *A Sourcebook of Problems for Geometry Based Upon Industrial Design and Architectural Ornament.* Mabel Sykes. Dale Seymour Pub, 364 pp, (P). [ISBN 0-86651-795-2] Brief historical comments and numerous exercises based on design analyses of many lovely geometric patterns. With updated bibliography; first published in 1912. JNC

Algebraic Topology, P. *Cobordisms and Spectral Sequences.* V.V. Vershinin. Transl. of Math. Mono., V. 130. AMS, 1993, v + 97 pp, \$62. [ISBN 0-8218-4582-9]

Algebraic Topology, P. *Topology and Representation Theory.* Eds: Eric M. Friedlander, Mark E. Mahowald. Contemp. Math., V. 158. AMS, 1994, ix + 318 pp, \$48 (P). [ISBN 0-8218-5165-9] Proceedings of a 1992 conference at Northwestern University.

Algebraic Topology, T*(17: 1), L. Homology Theory: An Introduction to Algebraic Topology, Second Edition. James W. Vick. Grad. Texts in Math., V. 145. Springer-Verlag, 1994, xiv + 242 pp, \$49. [ISBN 0-387-94126-6] New: an extensive chapter on covering spaces, presenting the lifting problem and using it to introduce the fundamental group. Stresses intuitive treatment of basic ideas. (First Edition, TR, August–September 1973.) DP

Operations Research, T(15–17: 1–3), C. Operations Research: Applications and Algorithms, Third Edition. Wayne L. Winston. Duxbury Pr, 1994, xx + 1372 pp, with disk. [ISBN 0-534-20971-8] Comprehensive introductory text; covers linear programming and other deterministic and probabilistic models. New edition has 200+ new problems, 6 case studies, rewritten chapter on nonlinear programming, and more emphasis on software (especially structured modeling packages). Disk contains student editions of LINDO, GINO, and LINGO as well as data files. (Second Edition, TR, April 1991.) AO

Control Theory, P. Lecture Notes in Control and Information Sciences-191: Simultaneous Stabilization of Linear Systems. Vincent Blondel. Springer-Verlag, 1994, xxi + 184 pp, \$45 (P). [ISBN 0-387-19862-8]

Control Theory, P. Lecture Notes in Control and Information Sciences-193: Variable Structure and Lyapunov Control. Ed: Alan S.I. Zinober. Springer-Verlag, 1994, xxii + 401 pp, \$78 (P). [ISBN 0-387-19869-5] 18 papers on theory and applications.

Control Theory, P. Lecture Notes in Control and Information Sciences-192: The Modeling of Uncertainty in Control Systems. Eds: Roy S. Smith, Mohammed Dahleh. Springer-Verlag, 1994, xv + 391 pp, \$79 (P). [ISBN 0-387-19870-9] 31 essays and technical papers from a 1992 workshop at the University of California, Santa Barbara.

Control Theory, P. Lecture Notes in Control and Information Sciences-188: Aeroassisted Orbital Transfer: Guidance and Control Strategies. D. Subbaram Naidu. Springer-Verlag, 1994, xiii + 178 pp, \$44 (P). [ISBN 0-387-19819-9]

Systems Theory, T(17: 1). The Elements of System Design. Amer A. Hassan, et al. Academic Pr, 1994, x + 280 pp, \$54.95. [ISBN 0-12-343060-7] Practical approach to designing large-scale digital systems. Two premises: (1) engineering education slights tradeoffs possible between individual models and off-the-shelf components; (2) systems design should

integrate 4 main disciplines: signal processing, communications, control, computation. RM

Stochastic Processes, P. Sojourn Times in Feedback and Processor Sharing Queues. J.L. van den Berg. CWI Tracts, V. 97. Centrum voor Wiskunde en Informatica, 1993, 123 pp, Dfl. 40 (P). [ISBN 90-6196-427-X]

Elementary Statistics, T(13: 1). Statistics: An Introduction, Fourth Edition. Robert D. Mason, Douglas A. Lind, William G. Marchal. Saunders College, 1994, xvii + 790 pp, \$53.25. [ISBN 0-03-096917-4] Traditional topics. Over 40% of text devoted to one-dimensional descriptive statistics and probability, 20% to basic statistical inference, the rest to one-way analysis of variance, simple and multiple regression and correlation, chi-square tests, nonparametric methods. RSK

Statistics, P. Stochastic Integrals and Goodness-of-fit-Tests. A.J. Koning. CWI Tracts, V. 98. Centrum voor Wiskunde en Informatica, 1993, iv + 163 pp, Dfl. 50 (P). [ISBN 90-6196-428-8]

Statistics, P. Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informal Approach, Volumes 1–3. Ed: H. Bozdogan. Kluwer Academic, 1994, \$390 set [ISBN 0-7923-2600-8]. Volume 1: Theory and Methodology of Time Series Analysis, xv + 277 pp; Volume 2: Multivariate Statistical Modeling, xiii + 413 pp; Volume 3: Engineering and Scientific Applications, xiii + 346 pp. 42 papers from a 1992 conference in Knoxville, Tennessee.

Statistics, P. The Chronological Annotated Bibliography of Order Statistics, Volume VIII: Indices, with a Supplement on 1970–1992. H. Leon Harter, N. Balakrishnan. Ser. in Math. & Management Sci., V. 24. American Sciences Pr, 1993, vii + 267 pp, \$110 (P). [ISBN 0-935950-26-5]

Programming, C. ObjectWindows for C++. Robert J. Traister. Academic Pr, 1993, xii + 205 pp, \$39.95 (P), with disk. [ISBN 0-12-697415-2] User-friendly introduction to creating Windows applications that allow text-editing, mouse operations, graphics. MPR

Computer Systems, P. SCO UNIX in a Nutshell: A Desktop Quick Reference for SCO UNIX & Open Desktop. Ellis Cutler. O'Reilly & Assoc, 1994, xix + 568 pp, \$9.95 (P). [ISBN 1-56592-037-6]

Theory of Computation, P. The Graph Isomorphism Problem: Its Structural Complexity. Johannes Köbler, Uwe Schöning, Jacobo Torán. Prog. in Theoret. Comp. Sci. Birkhäuser, 1993, 160 pp, \$34.50. [ISBN 0-8176-3680-3]

Computer Science, P. *Proceedings of the Fifth Annual ACM-SIAM Symposium on Discrete Algorithms.* ACM & SIAM, 1994, 735 pp, (P). [ISBN 0-89871-329-3]

Computer Science, P. *Proceedings: 8th Annual X Technical Conference.* Ed: Adrian Nye. The X Resource: A Practical Journal of the X Window System (Issue 9). O'Reilly & Assoc, 1994, 253 pp, \$22.50 (P). [ISBN 1-56592-066-X]

Computer Science, T*(12), C, L. *Computer Science: An Overview, Fourth Edition.* J. Glenn Brookshear. Ser. in Comp. Sci. Benjamin/Cummings, 1994, xii + 506 pp, \$40.50 (P). [ISBN 0-8053-4627-9] Surveys the "big picture:" machine architecture, software, data organization, future directions. New topics include object-oriented programming and databases, networks, the OSI reference model. Lab manuals in Pascal and C. (1985 text, TR, November 1987.) MPR

Computer Science, T(14-15), S, C, L. *The Mathematica Programmer.* Roman E. Maeder. Academic Pr, 1994, xv + 199 pp, \$44.95 (P), with disk. [ISBN 0-12-464990-4] An introductory computer science text with an emphasis on object-oriented programming concepts. Also an introduction to Mathematica as a programming language. Therein lies a problem—the intended audience is not clear. Nonetheless, an interesting book. MPR

Computer Science, P. *The CWEB System of Structured Documentation, Version 3.0.* Donald E. Knuth, Silvio Levy. Addison-Wesley, 1994, 226 pp, (P). [ISBN 0-201-57569-8] User's guide and reference manual for CWEB, a version of Knuth's WEB system for "literate programming" adapted to C and C++. AO

Computer Science, S(16-17), P*, L.** *The Stanford GraphBase: A Platform for Combinatorial Computing.* Donald E. Knuth. Addison-Wesley, 1993, viii + 575 pp. [ISBN 0-201-54275-7] A collection of "literate programs" that present important algorithms and data structures (e.g., Dijkstra's algorithm for shortest paths, Fibonacci heaps). Includes data for benchmarking combinatorial algorithms. AO

Applications (Biological Science), P. *Disease Dynamics.* Alexander Asachenkov, et al. Systems & Control: Found. & Applic. Birkhäuser, 1994, xv + 316 pp, \$79.50. [ISBN 0-8176-3692-7] Summarizes 16 years of research in theoretical immunology. A unified approach to the developing and using mathematical models of immune response. AO

Applications (Economics), P. *Computational Techniques for Econometrics and Economic Analysis.* Ed: D.A. Belsley. Adv. in Com-

putat. Econ., V. 3. Kluwer Academic, 1994, ix + 238 pp, \$92. [ISBN 0-7923-2356-4] 13 recent studies on uses of numerical techniques in economics and econometrics.

Applications (Fluid Dynamics), T(17-18: 1), P. *Multiphase Flow and Fluidization: Continuum and Kinetic Theory Descriptions.* Dimitri Gidaspow. Academic Pr, 1994, xx + 467 pp, \$69.50. [ISBN 0-12-282470-9] State-of-the-art treatment of multiphase flows in such engineering situations as boiling, condensation, separation, and mixing. Aims to help reader understand and use computer tools newly available to handle previously intractable problems. DS

Applications (Fluid Dynamics), P. *Annual Review of Fluid Mechanics, Volume 26, 1994.* Eds: John L. Lumley, Milton Van Dyke, Helen L. Reed. Annual Reviews, 1994, x + 704 pp, \$47. [ISBN 0-8243-0726-7]

Applications (Physics), P. *Deformations of Mathematical Structures II.* Ed: Julian Ławrynowicz. Kluwer Academic, 1994, x + 464 pp, \$199. [ISBN 0-7923-2576-1] Papers on Hurwitz-type structures and applications to surface physics.

Applications (Physics), P. *Wave Propagation: Scattering Theory.* Ed: M. Sh. Birman. Transl. Ser. 2, V. 157. AMS, 1993, x + 256 pp, \$105. [ISBN 0-8218-7507-8] Papers treat wave propagation, scattering theory, and linear differential and pseudodifferential operators.

Applications (Physics), P. *Quantum Scattering Theory for Several Particle Systems.* L.D. Faddeev, S.P. Merkuriev. Math. Physics & Appl. Math., V. 11. Kluwer Academic, 1993, xiii + 404 pp, \$180. [ISBN 0-7923-2414-5]

Applications (Physics), P. *Scattering in Quantum Field Theories: The Axiomatic and Constructive Approaches.* Daniel Jaegle. Ser. in Physics. Princeton Univ Pr, 1993, xxi + 290 pp, \$49.50. [ISBN 0-691-08589-7]

Applications (Physics), P. *Operator Algebras, Mathematical Physics, and Low Dimensional Topology.* Eds: Richard Herman, Betül Tanbay. Res. Notes in Math., V. 5. AK Peters, 1993, 324 pp, \$39.95. [ISBN 1-56881-027-X] Contributed papers from a 1991 NATO Advanced Research Workshop in Istanbul, Turkey.

Reviewers

JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; HD: Hung Dinh, Macalester; TH: Tom Halverson, Macalester; SK: Steve Kennedy, St. Olaf; RSK: Richard S. Kleber, St. Olaf; RM: Richard Molnar, Macalester; AO: Arnold Ostebee, St. Olaf; DP: David Peifer, St. Olaf; MPR: Matthew P. Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; KS: Karen Saxe, Macalester; JS: John Schue, Macalester; DS: Dan Schwalbe, Macalester; KES: Kay E. Smith, St. Olaf; MW: Martha Wallace, St. Olaf.

NEW IN THE SPECTRUM SERIES

The Words of Mathematics

An Etymological Dictionary of Mathematical Terms Used in English

Steven Schwartzman

Heteroscedastic *Amphicheiral* *Centroid* *Clothoid* *Eigenvalue*

The Words of Mathematics explains the origins of over 1500 mathematical terms used in English. While other dictionaries of mathematics define technical terms, this book concentrates on where those terms came from and what their literal meanings are. The words included here range from simple to advanced. Elementary school teachers may be surprised to learn that *inch* and *ounce* are really the same word, that *eleven* means literally "one left over," that a *thousand* is a "swollen hundred," and that the original meaning of *times* was "*divide*." High school teachers will find out that *asymptote* means "not falling together," that an *area* used to be a threshing floor, and that *focus* is a Latin word meaning "fire-place." College teachers who want to explain *heteroscedastic*, *amphicheiral*, and *eigenvalue* to their students will find the origins of those words in this book.

This dictionary is easy to use. Although some of the entries are highly technical, the book explains them in plain English. The introduction gives an overview of how the ancient language known as Indo-European developed into Latin, Greek, French, and English, the languages from which most of our mathematical vocabulary has been derived. Another section discusses the many ways

in which mathematicians have borrowed and created their specialized vocabulary over the centuries. A glossary explains historical and linguistic terms used throughout the book.

As in any dictionary, the entries themselves are arranged alphabetically. The words are drawn from arithmetic, algebra, geometry, trigonometry, calculus, number theory, topology, statistics, graph theory, logic, recreational mathematics, and other areas. Over 200 illustrations accompany the dictionary entries, especially some of the less familiar ones. Connections to related nonmathematical English words are often pointed out. Key numbers attached to many entries lead interested readers to an appendix which groups mathematical terms that come from a common source.

This dictionary is an indispensable reference for every library that serves teachers and students of mathematics. It is a natural source of information for courses in the history of mathematics and for mathematics courses intended for liberal arts students.

262 pp., Paperbound, 1994

ISBN 0-88385-511-9

List: \$27.00 MAA Member: \$21.00

Catalog Number WORDS

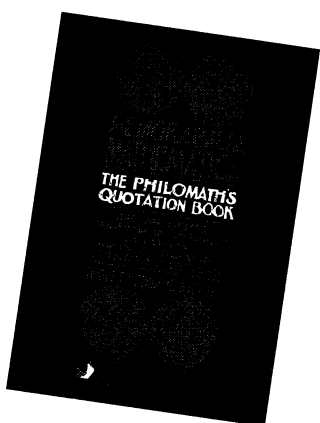
Name _____
Address _____
City _____
State _____ Zip Code _____

Qty.	Catalog Number	Price
		Total \$ _____
Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
Credit Card No. _____		
Signature _____		
Exp. Date _____		

Memorabilia Mathematica

The Philomath's Quotation Book

Robert Edouard Moritz



When Robert Edouard Moritz compiled his book of quotations, **Memorabilia Mathematica**, which appeared in 1914, he stated that his primary objective was to seek out the exact statement of and exact references for famous passages about mathematics. He searched the writing not only of mathematicians, but poets, philosophers, historians, statesmen, and scientists as well. His sources ranged from the works of Plato to the writings of Hilbert and Whitehead. His second objective was to produce a volume that would be a source of pleasure, encouragement, and inspiration to both mathematicians and non-mathematicians alike.

This work was a ten-year labor of love, and it is a tribute to his discerning eye that this selection of passages should remain one of the most stimulating works about mathematics ever published. It was the first collection of its kind in English and it conveys a sense of the full range of mathematics, its enormous accomplishments, and the living personalities of great mathematicians.

The more than eleven-hundred fully annotated selections in this book, gathered from the works of three hundred authors, cover a vast range of subjects pertaining to mathematics. Grouped in twenty-one chapters, they deal with such topics as the definitions and objects of mathematics; the teaching of mathematics; mathematics as a language or as a fine art; the relationship of mathematics to philosophy, to logic, or to science; the

nature of mathematics, and the value of mathematics. Other sections contain passages referring to specific subjects in the field such as arithmetic, algebra, geometry, calculus, and modern mathematics. Of special interest is the extensive amount of material on great mathematicians which provides irreplaceable glimpses into the lives and personalities of mathematical giants.

To mathematicians the book will be a great source of pleasure, inspiration, and encouragement. To teachers of mathematics and writers about mathematics, it will remain of inestimable value as a source of quotations and ideas. To the layperson, it will be a revelation. It should dispel forever the narrow notion that mathematics is a cut-and-dried affair, isolated from other compartments of life and thought.

440 pp., Paperbound, 1993
ISBN 0-88385-321-3
List: \$24.00 MAA Member: \$19.00
Catalog Number: MEMO

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Name _____

Address _____

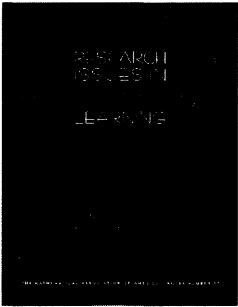
City _____

State _____ Zip Code _____

Qty.	Catalog Number	Price
_____	_____	_____
_____	_____	_____
		Total \$ _____
Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
Credit Card No. _____		
Signature _____		
Exp. Date _____		

Research Issues in Undergraduate Mathematics Learning Preliminary Analyses and Reports

James J. Kaput and Ed Dubinsky, Editors



Research in undergraduate mathematics education is important for all college and university mathematicians. If our students are to be more successful in understanding mathematics, then college faculty need to understand how mathematics is learned. This knowledge can guide us in curriculum reform and in improving our own teaching. It can help us make mathematics accessible to all students and it can increase the number of graduate students in mathematics.

This volume of research in undergraduate mathematics education informs us about the nature of student learning in some of the most important topics in the undergraduate curriculum: sets, functions, calculus, statistics, abstract algebra and problem solving. Paying careful attention to the trouble students have in learning mathematics will help us to work with students so they can deal with those difficulties.

A survey of the literature begins the volume. Becker and Pence have brought together an unusually complete list of references on research in collegiate mathematics. Their comments will guide those attempting to begin or to continue a program of research in student learning.

The sad fact that even good calculus students stumble over nonroutine problems is the theme of Selden, Selden, and Mason. Their conclusions point to significant shortcomings in the curriculum. This study of student difficulties is

continued by Ferrini-Mundy and Graham who investigate a single student's interactions with the fundamental concepts of the calculus. Baxter studies a group of students to learn how they acquire the concept of set, while Cuoco does the same for the concept of function.

Cooperative learning does help the student. That is the conclusion of Bonsangue, who investigates how two carefully matched classes of students in a statistics course perform on exams. How students learn to write proofs in group theory is the subject considered by Hart. Rosamond breaks new ground by comparing how emotions vary in their effect on the problem solving ability of novices and experts.

All college faculty should read this book to find how they can help their students learn mathematics.

150 pp., Paperbound, 1994
ISBN 0-88385-090-7

List: \$24.00
Catalog Number NTE-33

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Membership Code	Qty.	Catalog Number	Price
_____	_____	_____	_____
Name _____	Total \$ _____		
Address _____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
City _____	Credit Card No. _____		
State _____ Zip Code _____	Signature _____		
	Exp. Date _____		

FOUR TEXTS

I think that in retirement I have become a publisher. I am proud to offer a new text by a master of the field

Donald Sarason, *Notes on Complex Function Theory* (list price \$17), intended for a first, undergraduate course of one semester. This book fills a need for a straightforward, correct, elementary presentation.

I continue to offer three books by myself: *Linear Algebra* (first published by Holden-Day) (list price \$20), which presents the standard syllabus simply enough for ordinary junior-level courses, but without compromising mathematical integrity; and

Honors Calculus (list price \$24), addressed to freshmen who want to do calculus right the first time; and

Harmonic Analysis (first published by Addison-Wesley, reprinted by Wadsworth) (list price \$24), a graduate-level introduction to Fourier Series and the wonderful part of analysis that depends on this venerable subject.

Send orders or requests for examination copies to Henry Helson, 15 The Crescent, Berkeley, CA 94708; Tel (510)848-8629; email helson@math.berkeley.edu

Visualization in Teaching and Learning Mathematics

**Walter Zimmermann and
Steve Cunningham, Editors**

Buy this book. If you can't buy it, have the library order it. If the library won't order it, ask to borrow a copy from a friend. But do read this book.

—*The Mathematics Teacher*

High school, community college, and university teachers who use or are interested in using graphics to teach calculus, deductive reasoning, functions, geometry, or statistics will find valuable ideas for teaching... A must for every college or university library with a mathematics department.—CHOICE

The twenty papers in this book give an overview of research, analysis, practical experience, and informed opinion about the role of visualization in teaching and learning mathematics, especially at the undergraduate level. Visualization in its broadest sense is as old as mathematics, but

progress in computer graphics has generated a renaissance of interest in visual representations and visual thinking in mathematics.

230 pp., Paperbound, 1991

ISBN 0-88385-071-0

List: \$24.00

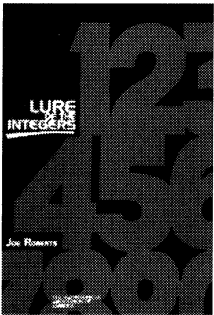
Catalog Number NTE-19

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

LURE OF THE INTEGERS

Joe Roberts



A joy to read and ponder, this book is a welcome addition to the body of mathematical literature. It belongs in every mathematical library.

—*Journal of Recreational Mathematics*

Will enrich library collections serving curricula with theory of numbers courses.

—*Choice*

In some small way, this book is an introduction to a mythical book which might go under the name of *The Book of Integers*. This mythical book has on page n all of the interesting properties of the integer n . This introduction stems from many years' casual accumulation of numerical facts. Most of the material presented belongs to elementary mathematics in the sense that no deep or profound mathematical background is required in order to understand what is said. Much of the material is drawn from the theory of numbers.

Many of the topics touch on contemporary research and most of the results are stated without proof. As a general rule, one cannot tell from the statements of the results whether or not their proofs will be elementary. Indeed, this is a hallmark of mathematics and is one of the things that gives the subject a special flavor and interest. Until one knows that expert practitioners have been unable to solve a problem, one does not know that the problem is difficult. Even then it may turn out that there is an easy solution.

Some of the material will be familiar to people having only a small acquaintance with mathematics. Even in those cases, the author provides something new. On the other hand, much of the material is sufficiently out of the main stream of concern that even professional mathematicians may be unfamiliar with the results. The many references to the literature will almost always enable a reader to track down further information. In **Lure of the Integers** the author has presented a body of material which will prove interesting to the enlightened layman as well as to the professional.

300 pp., Paperbound, 1992

ISBN-0-88385-502-X

List: \$31.50 MAA Member: \$21.50

Catalog Number LURE

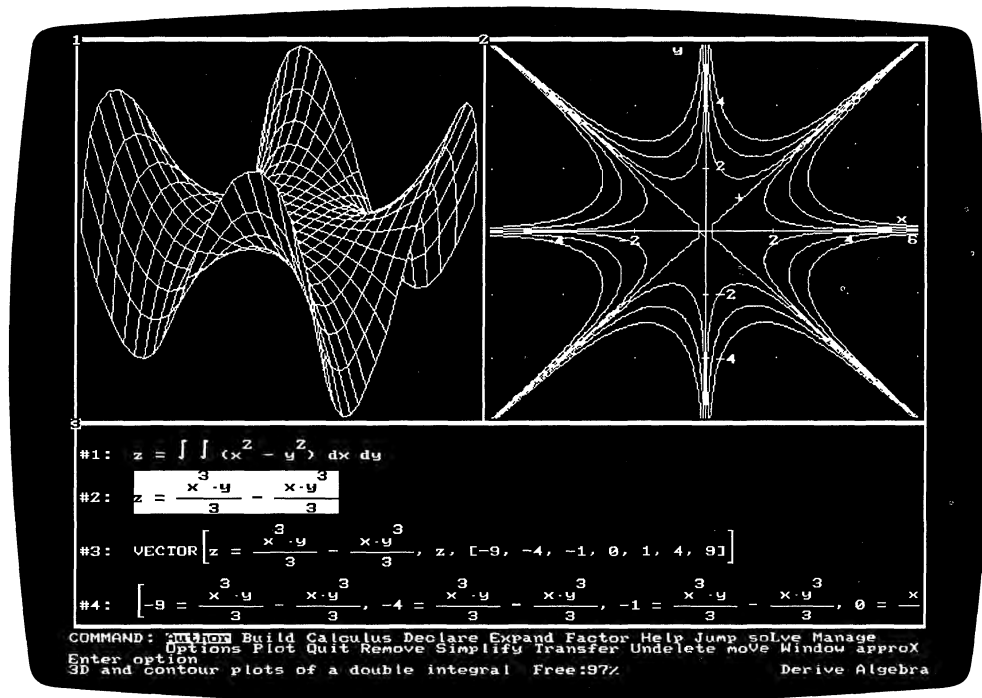
ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

Foreign Orders Please add \$3.00 per item ordered to cover postage and handling fees. The order will be sent via surface mail. If you want your order sent by air, we will be happy to send you a proforma invoice for your order.

Membership Code -----	Qty.	Catalog Number	Price
Name _____	_____		
Address _____	_____		
City _____	Total \$ _____		
State _____ Zip Code _____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
	Credit Card No. _____		
	Signature _____ Exp. Date _____		

NEW ~~DERIVE~~ VERSION 3!



DERIVE is a powerful computer algebra system for doing symbolic and numeric mathematics on your personal computer.

DERIVE:

- Performs numerical operations exactly with no round-off error
- Approximates irrational expressions to thousands of digits of precision
- Algebraically simplifies, expands, and factors expressions; and solves equations
- Applies the rules of trigonometry, calculus, matrix algebra, and vector calculus
- Plots explicitly and implicitly defined functions in 2D with zooming and auto-scaling
- Generates 3D wire-frame function plots using hidden-line removal
- Displays and prints expressions using standard 2D mathematical notation
- Provides an easy to use, menu-driven interface with on-line help
- Is ideal for students, teachers, engineers, scientists, and mathematicians

DERIVE Requirements

(regular memory version):
A PC compatible running MS-DOS with 512K memory and a 3 1/2 inch (720K) diskette drive or an HP 95LX, 100LX or 200LX palmtop computer with 1M memory and connectivity pack for downloading. List \$125.

DERIVE XM Requirements

(extended memory version):
A 386, 486 or Pentium® based PC compatible running MS-DOS version 3.0 or later with at least 2M of extended memory and a 3 1/2 inch (1.4M) diskette drive. List \$250.

DERIVE is a registered trademark of Soft Warehouse, Inc.



Soft Warehouse
HONOLULU • HAWAII

Soft Warehouse, Inc. • 3660 Waiialae Ave.
Ste. 304 • Honolulu, HI, USA 96816-3236
Ph. (808) 734-5801 • Fax. (808) 735-1105

The Wohascum County Problem Book

George T. Gilbert, Mark Krusemeyer,
and Loren C. Larson

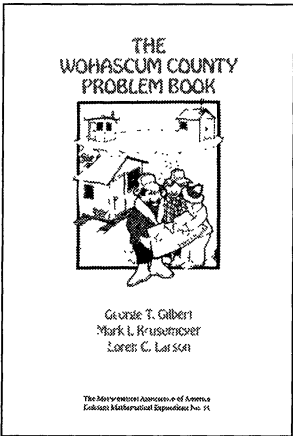
This collection is a delight. While the people and location are imaginary, their math ability is exceptionally real. A rich source of material for problem solving groups, preparation for competition or special credit courses, this book belongs in all academic libraries.
—Journal of Recreational Mathematics

This book consists of 130 problems that were originally presented as weekly challenges to undergraduate students. However, some of the problems could fall into the realm of secondary school mathematics. The book will make a welcome addition to any problem-solving fan's library.
—Alfred S. Posamentier, City College of the University of New York

If you like problem solving, this book belongs on your shelf.

Some knowledge of linear or abstract algebra is needed for a few of the problems, but most require nothing beyond calculus, and many should be accessible to high school students. However, there is a wide range of difficulty, and some problems require considerable mathematical maturity. For most students, few, if any, of the problems will be routine.

The book centers on solutions which are elegant, instructive, and clear. Often several solutions to the same problem are presented. Some problems have complicated solutions, and many of them



are quite long. Sometimes solutions are preceded by “Ideas,” which can serve as motivation or as hints, or followed by “Comments,” which often put solutions in a broader perspective.

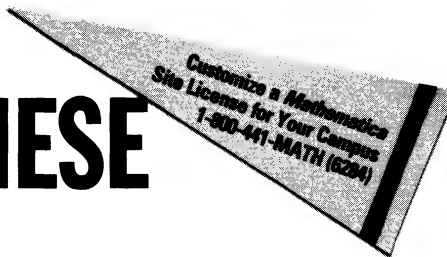
Indices are provided which may be especially helpful to problem solving classes and to teams of individuals preparing for contests such as the Putnam exam.

244 pp., Paperbound, 1993
ISBN 0-88385-316-7
List: \$28.00 MAA Member: \$21.00
Catalog Number DOL-14

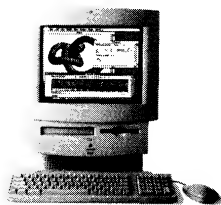
ORDER FROM:
The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

	Qty.	Catalog Number	Price
Membership Code -----			
Name _____			Total \$ _____
Address _____			Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
City _____			Credit Card No. _____
State _____ Zip Code _____			Signature _____
			Exp. Date _____

PUT ONE OF THESE



ON EVERY ONE OF THESE



FOR JUST ONE OF THESE.



How are universities around the world putting *Mathematica*® on every computer on campus for as little as \$1 per student? They are taking advantage of new *Mathematica* site



license programs. In fact, site licenses at over 700 universities have made *Mathematica* accessible to millions of students without breaking the school budget.

This new series of flexible, affordable site license programs

puts you in charge. You choose where you want *Mathematica*, what kinds of computers you want it on, how you want to network it in your labs, and how much you want to invest. And you can save up to 90% off the already-reduced academic price.



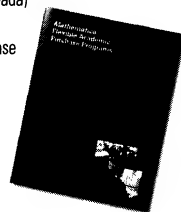
If you're excited about preparing more students for tomorrow by teaching with the world's leading technical computing system today, talk to us about a *Mathe-*

matica site license. We'll work together to customize one to fit your school's needs.

1-800-441-MATH (6284)

(U.S. and Canada)

Call us about a site license for your campus and ask for a free copy of this booklet, *Mathematica Flexible Academic Purchase Programs*.



Wolfram Research

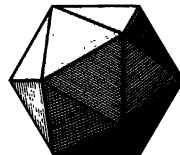
***Mathematica* is available for: Macintosh • Power Macintosh • Microsoft Windows • Microsoft Windows NT • MS-DOS • Sun SPARC • HP • Hitachi • DEC Alpha OSF/1, RISC, VAX/VMS • IBM RISC • SGI • NEC PC • NEC EWS • NEXTSTEP • CONVEX • and others.**

Corporate headquarters: **Wolfram Research, Inc.**, +1-217-398-0700; fax: +1-217-398-0747; email: info@wri.com. Europe: **Wolfram Research Europe Ltd.**, +44-(0)1993-883400; fax: +44-(0)1993-883800; email: info-euro@wri.com. Asia: **Wolfram Research Asia Ltd.** (Tokyo office), +81-(0)3-5276-0506; fax: +81-(0)3-5276-0509; email: info-asia@wri.com

© 1994 Wolfram Research, Inc. *Mathematica* is a registered trademark of Wolfram Research, Inc. *Mathematica* is not associated with Mathematica Policy Research, Inc. or MathTach, Inc. All other product names mentioned are trademarks of their producers

The American Mathematical Monthly

Volume 101 Number 9 / NOVEMBER 1994
(ISSN 0002-9890)



Contents

ARTICLES

- Georg Cantor and Transcendental Numbers / ROBERT GRAY 819
The 500th Anniversary of the Sharing Problem (The Oldest Problem
in the Theory of Probability) / MILTON SOBEL
and KRZYSZTOF FRANKOWSKI 833
What Is Teaching? / PAUL R. HALMOS 848
What Is Wrong with the Definition of dy/dx ? / HUGH THURSTON 855
A Stochastic Approach to the Gamma Function / LOUIS GORDON 858
Arrangements and Topological Planes / JACOB E. GOODMAN,
RICHARD POLLACK, RAPHAEL WENGER,
and TUDOR ZAMFIRESCU 866
An Application of Fourier Series to the Most Significant Digit Problem /
JEFF BOYLE 879

FEATURES

COMMENTS 818

NOTES

- Cross Product Identities in Arbitrary Dimension /
ANDREW DITTMER 887
A Non-Constant, Continuous Function on the Plane Whose Integral
on Every Line Is Zero / D. H. ARMITAGE 892
Chu's 1303 Identity Implies Bombieri's 1990 Norm-Inequality
(Via an Identity of Beauzamy and Dégot) /
DORON ZEILBERGER 894

THE COMPUTER SCIENCE SAMPLER

- How to Stay Competitive / CATHERINE C. McGEOCH 897

THE EVOLUTION OF ...

- On the Calculus of Variations and Its Major Influences
on the Mathematics of the First Half of Our Century. Part II /
ERWIN KREYSZIG 902

THE AUTHORS 909

PROBLEMS AND SOLUTIONS 911

REVIEWS

- Fleeting Footsteps: Tracing the Conception of Arithmetic and Algebra
in Ancient China.* By Lam Lay Yong and Ang Tian Se /
FRANK SWETZ 921

TELEGRAPHIC REVIEWS 924

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.



INDEX TO VOLUME 101, 1994
THE AMERICAN MATHEMATICAL MONTHLY

TITLE INDEX

- Introduction to Fermat's Last Theorem, D. Cox, 3
- Future Elementary Teachers: The Neglected Constituency, T. W. Hungerford, 15
- Galois Theory for Beginners, J. Stillwell, 22
- Into the Hourglass: Reflections on the Forces Acting on a Granular Material, E. B. Pitman, 28
- Orderly Currencies, J. D. Jones, 36
- An Interior Fixed Point Property of the Disc, R. F. Brown and R. E. Greene, 39
- Le Cam's Inequality and Poisson Approximations, J. M. Steele, 48
- Turing Machines and Computational Complexity, B. Marion, 61
- The Evolution of Integration, A. Shenitzer and J. Stepāns, 66
- Yueh-Gin Gung and Dr. Charles Y. Hu Award for Distinguished Service to J. Sutherland Frame, D. W. Ballew, 107
- A New Look at Euler's Theorem for Polyhedra, B. Grünbaum and G. C. Shephard, 109
- Otto Neugebauer: Reminiscences and Appreciation, P. J. Davis, 129
- From the Buffon Needle Problem to the Kreiss Matrix Theorem, E. Wegert and L. N. Trefethen, 132
- A Counterexample for Germain, W. C. Waterhouse, 140
- Cubic Equations, or Where Did the Examination Question Come From?, H. B. Griffiths and A. E. Hirst, 151
- Every Number is Expressible as the Sum of How Many Polygonal Numbers?, R. K. Guy, 169
- A Marvelous Proof, F. Q. Gouvêa, 203
- Triangulating the Circle, at Random, D. Aldous, 223
- Hypatia and Her Mathematics, M. A. B. Deakin, 234
- Calculus II and Euler Also (with a Nod to Series Integral Remainder Bounds), R. Barshinger, 244
- A Focusing Property of the Ellipse, M. Frantz, 250
- Universal Traversal Sequences, J. Feigenbaum and N. Reingold, 262
- What are Algebraic Integers and What Are They For?, J. Stillwell, 266
- Pizza Slicing, Phi's and the Riemann Hypothesis, E. A. Bender, O. Patashnik, and H. Rumsey, Jr., 307
- Rational Periodic Points of the Quadratic Function $Q_c(x)=x^2+c$, R. Walde and P. Russo, 318
- Fréchet vs. Carathéodory, E. Acosta and C. Delgado, 332
- Odd Magic Powers, A. C. Thompson, 339
- Mathematicians, Including Undergraduates, Look at Soap Bubbles, F. Morgan, 343
- ApSimon's Mints Problem, R. Guy and R. Nowakowski, 358
- On the Geometry of Piecewise Circular Curves, T. Banchoff and P. Giblin, 403
- The Two Envelope Paradox, E. Linzer, 417
- Fourier Series of Polygons, A. Robert, 420
- The Paradox of Nontransitive Dice, R. P. Savage, Jr., 429
- Squares Expressible as Sum of Consecutive Squares, L. Beeckmans, 437
- Square Roots mod p , S. M. Turner, 443
- Does Anybody Really Know What Time It Is?, C. C. McGeoch, 459
- How Hyperbolic Geometry Became Respectable, A. Shenitzer, 464
- Juggling Drops and Descents, J. Buhler, D. Eisenbud, R. Graham, and C. Wright, 507
- Teaching Integration by Substitution, D. Gale, 520
- Workable Gears, Archimedian Solids and Planar Bipartite Graphs, G. Gordon, 527
- On the Kummer Solutions of the Hypergeometric Equation, R. T. Prosser, 535
- Reflections on a Mira, J. W. Emert, K. I. Meeks, and R. B. Nelson, 544
- Buffon Noodles, E. Waymire, 550
- A Possible Permanent Formula, D. Callan, 571
- A Tale of Two CD's, D. Kennedy, 603
- Three Problems in Search of a Measure, J. L. King, 609
- The n -Queens Problem, I. Rivin, I. Vardi, and P. Zimmermann, 629
- What's the Difference Between Cantor Sets?, R. L. Kraft, 640
- Morphisms, Squarefree Strings, and the Tower of Hanoi Puzzle, J.-P. Allouche, D. Astoorian, J. Randall, and J. Shallit, 651
- Do You Know the Way to Vertex A ?, J. Ondich, 668
- On the Calculus of Variations and Its Major

- Influences on the Mathematics of the First Half of Our Century. Part I., E. Kreyszig, 674
- Behind the Scenes of a Random Dot Stereogram, M. S. Terrell and R. E. Terrell, 715
- The Fifty-Fourth William Lowell Putnam Mathematical Competition, L. F. Klosinski, G. L. Alexanderson, and L. C. Larson, 725
- Literacy in the Language of Mathematics, J. O. Bullock, 735
- Fractional and Trigonometric Expressions for Matrices, G. Shimura, 744
- Noether Lasker Primary Decomposition Revisited, B. L. Osofsky, 759
- Elementary Infinite Sources of Non-Unique Factorization Rings, S. Stein and S. Szabó, 769
- Apropos* Two Notes on Notation, A. E. Fekete, 771
- Which Triangles are Plane Sections of Regular Tetrahedra?, F. Eriksson, 788
- Georg Cantor and Transcendental Numbers, R. Gray, 819
- The 500th Anniversary of the Sharing Problem (The Oldest Problem in the Theory of Probability), M. Sobel and K. Frankowski, 833
- What is Teaching?, P. R. Halmos, 848
- What is Wrong with the Definition of dy/dx ?, H. Thurston, 855
- A Stochastic Approach to the Gamma Function, L. Gordon, 858
- Arrangements and Topological Planes, J. E. Goodman, R. Pollack, R. Wenger, and T. Zamfirescu, 866
- An Application of Fourier Series to the Most Significant Digit Problem, J. Boyle, 879
- How to Stay Competitive, C. C. McGeoch, 897
- On the Calculus of Variations and Its Major Influences on the Mathematics of the First Half of Our Century. Part II., E. Kreyszig, 902
- The Rectilinear Crossing Number of a Complete Graph and Sylvester's "Four Point Problem" of Geometric Probability, E. R. Scheinerman and H. S. Wilf, 939
- String Matching for the Novice, O. E. Percus and J. K. Percus, 944
- Bernoulli Trials and Number Theory, D. Rawlings, 948
- What is the Shape of a Mylar Balloon?, W. H. Paulsen, 953
- Euler's Theorem for Polyhedra: A Topologist and Geometer Respond, P. Hilton and J. Pedersen, 959
- The Role of Paradoxes in the Evolution of Mathematics, I. Kleiner and N. Movshovitz-Hadar, 963
- Regions in the Complex Plane Containing the Eigenvalues of a Matrix, R. A. Brualdi and S. Mellendorf, 975
- Characterization of Solvable Quintics x^5+ax+b , B. K. Spearman and K. S. Williams, 986
- A Halmos Problem and a Related Problem, J. B. Cosgrave, 993
- Mousetrap, R. K. Guy and R. J. Nowakowski, 1007

AUTHOR INDEX

- Acosta, Ernesto and Cesar Delgado G., Fréchet vs. Carathéodory, 332
- Aldous, David, Triangulating the Circle, at Random, 223
- Alexanderson, Gerald L. *see Klosinski*
- Allouche, Jean-Paul, Dan Astoorian, Jim Randall, and Jeffrey Shallit, Morphisms, Squarefree Strings, and the Tower of Astoorian, Dan *see Allouche*
- Ballew, David W., Yueh-Gin Gung and Dr. Charles Y. Hu Award for Distinguished Service to J. Sutherland Frame, 107
- Banchoff, Thomas and Peter Giblin, On the Geometry of Piecewise Circular Curves, 403
- Barshinger, Richard, Calculus II and Euler Also (with a Nod to Series Integral Remainder Bounds), 244
- Beeckmans, Laurent, Squares Expressible as Sum of Consecutive Squares, 437
- Bender, Edward A., Oren Patashnik, and Howard Rumsey, Jr., Pizza Slicing, Phi's and the Riemann Hypothesis, 307
- Boyle, Jeff, An Application of Fourier Series to the Most Significant Digit Problem, 879
- Brown, Robert F. and Robert E. Greene, An Interior Fixed Point Property of the Disc, 39
- Brualdi, Richard A. and Stephen Mellendorf, Regions in the Complex Plane Containing the Eigenvalues of a Matrix, 975
- Buhler, Joe, David Eisenbud, Ron Graham, and Colin Wright, Juggling Drops and Descents, 507
- Bullock, James O., Literacy in the Language of Mathematics, 735
- Callan, David, A Possible Permanent Formula, 571
- Century. Part II., 902
- Century. Part I., 674
- Cosgrave, John B., A Halmos Problem and a Related Problem, 993
- Cox, David, Introduction to Fermat's Last Theorem, 3
- Davis, Philip J., Otto Neugebauer: Reminiscences and Appreciation, 129
- Deakin, Michael A. B., Hypatia and Her Mathematics, 234
- Delgado, Cesar *see Acosta*
- Eisenbud, David *see Buhler*
- Emert, John W., Kay I. Meeks, and Roger B. Nelson, Reflections on a Mira, 544
- Eriksson, Folke, Which Triangles are Plane Sections of Regular Tetrahedra?, 788
- Feigenbaum, Joan and Nick Reingold, Universal Traversal Sequences, 262
- Fekete, Antal E., *Apropos* Two Notes on Notation, 771
- Four Point Problem of Geometric Probability, 939
- Frankowski, Krzysztof *see Sobel*
- Frantz, Marc, A Focusing Property of the Ellipse, 250
- Gale, David, Teaching Integration by Substitution, 520
- Giblin, Peter *see Banchoff*
- Goodman, Jacob E., Richard Pollack, Rephael Wenger, and Tudor Zamfirescu, Arrangements and Topological Planes, Gordon, Louis, A Stochastic Approach to the Gamma Function, 858
- Gordon, Gary, Workable Gears, Archimedean Solids and Planar Bipartite Graphs, 527
- Gouvêa, Fernando Q., A Marvelous Proof, 203
- Graham, Ron *see Buhler*
- Gray, Robert, Georg Cantor and Transcendental Numbers, 819
- Greene, Robert E. *see Brown*
- Griffiths, H. B. and A. E. Hirst, Cubic Equations, or Where Did the Examination Question Come From?, 151
- Grünbaum, Branko and G. C. Shephard, A New Look at Euler's Theorem for Polyhedra, 109
- Guy, Richard and Richard Nowakowski, ApSimon's Mints Problem, 358
- Guy, Richard K. and Richard J. Nowakowski, Mousetrap, 1007
- Guy, Richard K., Every Number is Expressible as the Sum of How Many Polygonal Numbers?, 169
- Halmos, Paul R., What is Teaching?, 848
- Hanoi Puzzle, 651
- Hilton, Peter and Jean Pedersen, Euler's Theorem for Polyhedra: A Topologist and Geometer Respond, 959
- Hirst, A. E. *see Griffiths*
- Hungerford, Thomas W., Future Elementary Teachers: The Neglected Constituency, 15
- Jones, John Dewey, Orderly Currencies, 36
- Kennedy, Dan, A Tale of Two CD's, 603
- King, Jonathan L., Three Problems in Search of a Measure, 609
- Kleiner, I. and N. Movshovitz-Hadar, The Role of Paradoxes in the Evolution of Mathematics, 963
- Klosinski, Leonard F., Gerald L. Alexanderson, and Loren C. Larson, The Fifty-Fourth William Lowell Putnam Kraft, 866
- Roger L., What's the Difference Between Cantor Sets?, 640
- Kreyszig, Erwin, On the Calculus of Variations and Its Major Influences on the Mathematics of the First Half of Our Century, 640
- Kreyszig, Erwin, On the Calculus of Variations and Its Major

- Influences on the Mathematics of the First Half of Our Larson, Loren C. *see Klosinski*
- Linzer, Elliot, The Two Envelope Paradox, 417
- Marion, Bill, Turing Machines and Computational Complexity, 61
- Mathematical Competition, 725
- McGeoch, Catherine C., Does Anybody Really Know What Time It Is?, 459
- McGeoch, Catherine C., How to Stay Competitive, 897
- Meeks, Kay I. *see Emert*
- Mellendorf, Stephen *see Brualdi*
- Morgan, Frank, Mathematicians, Including Undergraduates, Look at Soap Bubbles, 343
- Movshovitz-Hadar, N. *see Kleiner*
- Nelson, Roger B. *see Emert*
- Nowakowski, Richard *see Guy*
- Nowakowski, Richard J. *see Guy*
- Ondich, Jeff, Do You Know the Way to Vertex A?, 668
- Osofsky, Barbara L., Noether Lasker Primary Decomposition Revisited, 759
- Patashnik, Oren *see Bender*
- Paulsen, William H., What is the Shape of a Mylar Balloon?, 953
- Pedersen, Jean *see Hilton*
- Percus, Ora E. and Jerome K. Percus, String Matching for the Novice, 944
- Percus, Jerome K. *see Percus*
- Pitman, E. Bruce, Into the Hourglass: Reflections on the Forces Acting on a Granular Material, 28
- Pollack, Richard *see Goodman*
- Prosser, Reese T., On the Kummer Solutions of the Hypergeometric Equation, 535
- Randall, Jim *see Allouche*
- Rawlings, Don, Bernoulli Trials and Number Theory, 948
- Reingold, Nick *see Feigenbaum*
- Rivin, Igor, Ilan Vardi, and Paul Zimmermann, The n -Queens Problem, 629
- Robert, Alain, Fourier Series of Polygons, 420
- Rumsey Jr., Howard, *see Bender*
- Russo, Paula *see Walde*
- Savage, Richard P., Jr., The Paradox of Non-transitive Dice, 429
- Scheinerman, Edward R. and Herbert S. Wilf, The Rectilinear Crossing Number of a Complete Graph and Sylvester's Shallit, Jeffrey *see Allouche*
- Shenitzer, A. and J. Steprāns, The Evolution of Integration, 66
- Shenitzer, Abe, How Hyperbolic Geometry Became Respectable, 464
- Shephard, G. C. *see Grünbaum*
- Shimura, Goro, Fractional and Trigonometric Expressions for Matrices, 744
- Sobel, Milton and Krzysztof Frankowski, The 500th Anniversary of the Sharing Problem (The Oldest Problem in the Spearman, Blair K. and Kenneth S. Williams, Characterization of Solvable Quintics x^5+ax+b , 986
- Steele, J. Michael, Le Cam's Inequality and Poisson Approximations, 48
- Stein, S. and S. Szabó, Elementary Infinite Sources of Non-Unique Factorization Rings, 769
- Steprāns, J. *see Shenitzer*
- Stillwell, John, Galois Theory for Beginners, 22
- Stillwell, John, What are Algebraic Integers and What Are They For?, 266
- Szabó, S. *see Stein*
- Terrell, Maria S. and Robert E. Terrell, Behind the Scenes of a Random Dot Stereogram, 715
- Terrell, Robert E. *see Terrell*
- Theory of Probability), 833
- Thompson, A. C., Odd Magic Powers, 339
- Thurston, Hugh, What is Wrong with the Definition of dy/dx ?, 855
- Trefethen, Lloyd N. *see Wegert*
- Turner, Stephen M., Square Roots mod p , 443
- Vardi, Ilan *see Rivin*
- Walde, Ralph and Paula Russo, Rational Periodic Points of the Quadratic Function $Q_c(x)=x^2+c$, 318
- Waterhouse, William C., A Counterexample for Germain, 140
- Waymire, Ed, Buffon Noodles, 550
- Wegert, Elias and Lloyd N. Trefethen, From the Buffon Needle Problem to the Kreiss Matrix Theorem, 132
- Wenger, Raphael *see Goodman*
- Wilf, Herbert S. *see Scheinerman*
- Williams, Kenneth S. *see Spearman*
- Wright, Colin *see Buhler*
- Zamfirescu, Tudor *see Goodman*
- Zimmermann, Paul *see Rivin*

NOTES TITLE INDEX

- A Generalization of a Theorem of Euler, Dorina Mitrea and Marius Mitrea, 55
- A Short Elementary Proof of the Mohr-Mascheroni Theorem, Norbert Hungerbühler, 784
- A Non-Constant, Continuous Function on the Plane Whose Integral on Every Line is Zero, D. H. Armitage, 892
- A "Popular" Class Number Formula, Kurt Girstmair, 997
- A Reverse Stolarsky's Inequality, Josip Pečarić, 565
- A Proof of Dilworth's Chain Decomposition Theorem, Fred Galvin, 352
- A Note on Some Irrational Decimal Fractions, A. McD. Mercer, 567
- A Trace Inequality for Unitary Matrices, Boying Wang and Fuzhen Zhang, 453
- An Elementary Proof of the Square Summability of the Discrete Hilbert Transform, Loukas Grafakos, 456
- Chaos Without Nonperiodicity, Carsten Knudsen, 563
- Chu's 1303 Identity Implies Bombieri's 1990 Norm-Inequality (Via an Identity of Beauzamy and Dégot), Doron Zeilberger, 894
- Congruence of Triangles, Leonard Gillman, 782
- Cross Product Identities in Arbitrary Dimension, Andrew Dittmer, 887
- Euler's Theorem, Katherine Heinrich and Peter Horak, 260
- Isometries of l_p -norm, Chi-Kwong Li and Wasin So, 452
- Kummer's Test Gives Characterizations for Convergence or Divergence of all Positive Series, Jingcheng Tong, 450
- More on the Pompeiu Problem, David C. Ulrich, 165
- New Tricks for Old Trees: Maps and the Pigeonhole Principle, N. Graham, R. C. Entringer, and L. A. Skékely, 664
- On Nonnegativity of Symmetric Polynomials, F. Matúš, 661
- On a Curious Property of Counting Sequences, Victor Bronstein and Aviezri S. Fraenkel, 560
- On Intervals, Transitivity = Chaos, Michel Vellekoop and Raoul Berglund, 353
- On the Identity of Polyhedra, Hellmuth Stachel, 162
- Proof of a Mixed Arithmetic-Mean, Geometric-Mean Inequality, Kiran Kedlaya, 355
- Reflections Can Be Trapped, Roberto Peirone, 259
- Sierpinski's Theorem is Deducible from Euler and Dirichlet, A. A. Ageev, 659
- The Existence of a Triangle with Prescribed Angle Bisector Lengths, Petru Mironescu and Laurentiu Panaitopol, 58
- The Coin Exchange Problem for Arithmetic Progressions, Amitabha Tripathi, 779
- The Second-Partials Test for Local Extrema of $f(x,y)$, Leonard Gillman, 1004
- Variations on Wolstenholme's Theorem, Emre Alkan, 1001

NOTES AUTHOR INDEX

- Ageev, A. A., Sierpinski's Theorem is Deducible from Euler and Dirichlet, 659
- Alkan, Emre, Variations on Wolstenholme's Theorem, 1001
- Armitage, D. H., A Non-Constant, Continuous Function on the Plane Whose Integral on Every Line is Zero, 892
- Berglund, Raoul *see Vellekoop*
- Bronstein, Victor and Aviezri S. Fraenkel, On a Curious Property of Counting Sequences, 560
- Dégot, 894
- Dittmer, Andrew, Cross Product Identities in Arbitrary Dimension, 887
- Entringer, R. C. *see Graham*
- Fraenkel, Aviezri S. *see Bronstein*
- Galvin, Fred, A Proof of Dilworth's Chain Decomposition Theorem, 352
- Gillman, Leonard, The Second-Partials Test for Local Extrema of $f(x,y)$, 1004
- Gillman, Leonard, Congruence of Triangles, 782
- Girstmair, Kurt, A "Popular" Class Number Formula, 997
- Grafakos, Loukas, An Elementary Proof of the Square Summability of the Discrete Hilbert Transform, 456
- Graham, N., R. C. Entringer, and L. A. Skékely, New Tricks for Old Trees: Maps and the

- Pigeonhole Principle, 664
 Heinrich, Katherine and Peter Horak, Euler's Theorem, 260
 Horak, Peter *see* Heinrich
 Hungerbühler, Norbert, A Short Elementary Proof of the Mohr-Mascheroni Theorem, 784
 Kedlaya, Kiran, Proof of a Mixed Arithmetic-Mean, Geometric-Mean Inequality, 355
 Knudsen, Carsten, Chaos Without Non-periodicity, 563
 Li, Chi-Kwong and Wasin So, Isometries of l_p -norm, 452
 Matúš, F., On Nonnegativity of Symmetric Polynomials, 661
 Mercer, A. McD., A Note on Some Irrational Decimal Fractions, 567
 Mironescu, Petru and Laurentiu Panaitopol, The Existence of a Triangle with Prescribed Angle Bisector Lengths, 58
 Mitrea, Marius *see* Mitrea
 Mitrea, Dorina and Marius Mitrea, A Generalization of a Theorem of Euler, 55
 Panaitopol, Laurentiu *see* Mironescu
 Pečarić, Josip, A Reverse Stolarsky's Inequality, 565
 Peirone, Roberto, Reflections Can Be Trapped, 259
 Skékely, L. A. *see* Graham
 So, Wasin *see* Li
 Stachel, Hellmuth, On the Identity of Polyhedra, 162
 Tong, Jingcheng, Kummer's Test Gives Characterizations for Convergence or Divergence of all Positive Series, 450
 Tripathi, Amitabha, The Coin Exchange Problem for Arithmetic Progressions, 779
 Ullrich, David C., More on the Pompeiu Problem, 165
 Vellekoop, Michel and Raoul Berglund, On Intervals, Transitivity = Chaos, 353
 Wang, Boying and Fuzhen Zhang, A Trace Inequality for Unitary Matrices, 453
 Zeilberger, Doron, Chu's 1303 Identity Implies Bombieri's 1990 Norm-Inequality (Via an Identity of Beuzamy and Zhang, Fuzhen *see* Wang

REVIEWS BY TITLE

Names of authors are in ordinary type; those of reviewers in capitals.

- Brouwer's Intuitionism*, Walter P. van Stigt, C. SMORYŃSKI, 799
Calculus Gems: Brief Lives and Memorable Mathematics, George F. Simmons, DAVID J. PENGELLEY, 374
Complex Analysis: The Geometric Viewpoint, Steven G. Krantz, JOHN POLKING, 91
Excursions in Calculus: An Interplay of the Continuous and the Discrete, Robert M. Young, ANITA E. SOLOW, 482
Fleeting Footsteps: Tracing the Conception of Arithmetic and Algebra in Ancient China, Lam Lay Yong and Ang Tian Se, FRANK SWETZ, 921
Geometry of Surfaces, John Stillwell, DAVID L. WEBB, 188
How to Teach Mathematics, Steven G. Krantz, MEYER JERISON, 692
Ideals, Varieties, and Algorithms, David Cox, John Little, and Donal O'Shea, MOSS SWEEDLER, 582
 Kolmogorov and A. P. Yushkevich, KAREN HUNGER PARSHALL, 369
Linear Programs and Related Problems, Evar D. Nering and Albert W. Tucker, STEPHEN B. MAURER, 1022
Mathematical Cranks, Underwood Dudley, IAN STEWART, 87
Mathematics of the 19th Century: Mathematical Logic, Algebra, Number Theory, Probability Theory, edited by A. N. Reality Rules I. *The Fundamentals*; II. *The Frontier*, John Casti, RUTHERFORD ARIS, 186
Revolutions in Mathematics, edited by Donald Gillies, MICHAEL MAHONEY, 283
The Lure of the Integers, Joe Roberts, PAUL T. BATEMAN and HAROLD G. DIAMOND, 480

SOLUTIONS

Numbers in boldface refer to problems; those in lightface to pages.

E3436	78	10185	85	10220	687	10245	797
E3464	81	10188	275	10221	796	10246	367
E3470	83	10195	277	10224	796	10247	581
E3474	177	10196	363	10225	181	10249	798
5881	794	10199	278	10227	688	10256	478
6080	913	10200*	1020	10228	183	10257	479
6332	1015	10203	279	10229	797	10261	690
6639	683	10204	365	10232	917	10263	1018
6649	77	10211	578	10235	280	10271	918
6657	80	10212	579	10237	366	10283	919
6666	576	10214	86	10239	281	10303	1019
6667	914	10215	580	10240	184		
6670*	1020	10216	917	10241	475		
6673	686	10218	179	10243	1017	*Revivals	
6674	179	10219	181	10244	476		

PROBLEMS PROPOSED

Abe, Nobuhisa 362	Hensley, Douglas <i>see Borosh</i>	Robinson, Raphael M. 76
Ali, Hassan Ali Shāh 75	Hirsch, Michael 363	Robinson, Raphael M. 574
Alkan, Emre 273	Hobbs, Arthur M. <i>see Borosh</i>	Robinson, Raphael M. 176
Alkan, Emre 574	Huanxin, Jiang 76	Rouhāni, Behzad Djafari 792
Andersen, E. Sparre 474	Kemp, Franklin 474	Santmyer, Joseph M. 175
Anglesio, Jean 575	King, Jonathan L. <i>see Hirsch</i>	Santos, Bernardo Recamán 273
Boese, F. G. 575	Kitchen, Edward 912	Satnoianu, Răzvan 1013
Book, David L. 274	Klamkin, Murray S. 575	Schindler, Werner 682
Borosh, Itshak 682	Knuth, Donald E. 682	Schmidt, Frank 911
Brillhart, John 362	Kotlarski, Ignacy I. 575	Schmidt, Frank 176
Chalice, Donald R. 176	Krompart, Lucia B. <i>see Fisher</i>	Selvaraj, S. <i>see Selvaraj</i>
Chen, Kwang-Wu 912	Kuczman, Marcin E. 363	Selvaraj, C. R. 1014
Chernoff, Paul R. 274	Laberteaux, Kathryn R. 362	Shor, Peter W. 793
Collins, Karen L. <i>see Fisher</i>	Larsen, Mogens Esrom <i>see Andersen</i>	Sieben, Nándor 474
Cornea, Emil A. 175	Liebeck, Hans 175	Simpson, R. J. 912
Correll Jr., Bill 1014	Lomont, J. S. <i>see Brillhart</i>	Sinnamon, Gord 911
Costa, Peter J. <i>see Rabinowitz</i>	Mathias, Roy 793	Slater, Michael 363
Darling, Donald A. 911	Militaru, Gigel 1014	Smyth, W. F. <i>see Simpson</i>
Diacu, Florin N. <i>see Cornea</i>	Mocanu, Mirel 912	Stanley, Richard P. 76
Doster, David 792	Nicol, C. <i>see Filaseta</i>	Stoyanov, Emil Yankov 274
Ekhad, Shalosh B. 75	Northshield, Sam 681	Tabov, Jordan 474
Enchev, Ognian 574	Osborne, Anthony <i>see Liebeck</i>	Tisdale, Daniel 176
Evans, Anthony <i>see Borosh</i>	Palacios, José Luis <i>see Northshield</i>	Tomescu, Ioan 681
Feldman, Jacob <i>see Chernoff</i>	Pelling, M. J. 274	van de Lune, J. 794
Fieldsteel, Adam 1014	Poonen, Bjorn 273	Vanden Eynden, Charles 1013
Filaseta, M. 1014	Poonen, Bjorn 363	
Fisher, David C. 793	Quet, Leroy 682	
Ford, Kevin 473	Rabinowitz, Stanley 474	
Gessel, Ira 75		
Gómez Rey, Joaquín 75		
Güllicher, Herbert 793		
Guy, Richard K. 473		
Hajja, Mowaffaq 682		

PROBLEMS SOLVED

Anchorage Math Solutions Group *see Chapman*

Andreoli, Michael 364

Barger, S. F. 798

Beckwith, David 80

Bromberg, Ken 476

Brulois, Frédéric 281

Călinescu, Gruia 276

Callan, David 181

Chakerian, G. D. 688

Chapman, Robin J. 184, 278, 478, 1017

Chen, William Y. C. 278

Chernoff, Paul R. 366

Darling, Donald A. 476

Ellingham, M. N. 80

Fasino, Dario 918

Firey, W. J. *see Chakerian*

Gagola Jr., Stephen M. 578

Georgiou, C. 475

Gregory, Michael B. 181

Griffin, Peter 78

Herman, Eugene A. 479, 914

Holzager, Richard 576, 688, 796, 917, 1019

Holzager, Richard *see Sagan*

Israel, Robert B. 183, 280, 687

Kastanas, Ilias 179, 578, 690

Klein, Benjamin G. 796

Knuth, D. E. 77

Komanda, Nasha 279

Krafft, Olaf 82

Laurie, Dirk P. 86

Lindsey II, John H. 683

Locke, Stephen C. 686

Lossers, O. P. *see Gagola*

Lossers, O. P. 79, 84, 797

Markham, Thomas L. 82

Mauldon, J. G. 1018

Merino, Dennis I. 82

Myerson, Gerry 85

NSA Problems Group 177

Pelling, M. J. 794, 913

Pinsky, Mark A. 690

Powers, R. Glenn *see Kastanas*

Rainwater, John 919

Royle, Gordon F. *see Ellingham*

Sagan, Bruce E. 686

Salinier, Alain 690

Seiffert, Heinz-Jürgen 690, 1015

Skau, Ivar 275

Steutel, F. W. 367

Stewart, Pat 580

Stong, Richard 180, 917

The Geometry Center *see Bromberg*

Timar, C. C. *see Ellingham*

Vowe, Michael 83

Wagon, Stan *see Bromberg*

Wells, David M. 365

Wertheim, Lev 581

Wilf, Herbert S. *see Locke*

Yiu, Paul 798

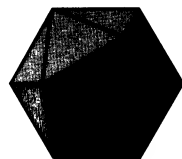
THANKS

The Monthly expresses its appreciation to the following people for their help in refereeing during the past year. We could not function without such people and their hard work.

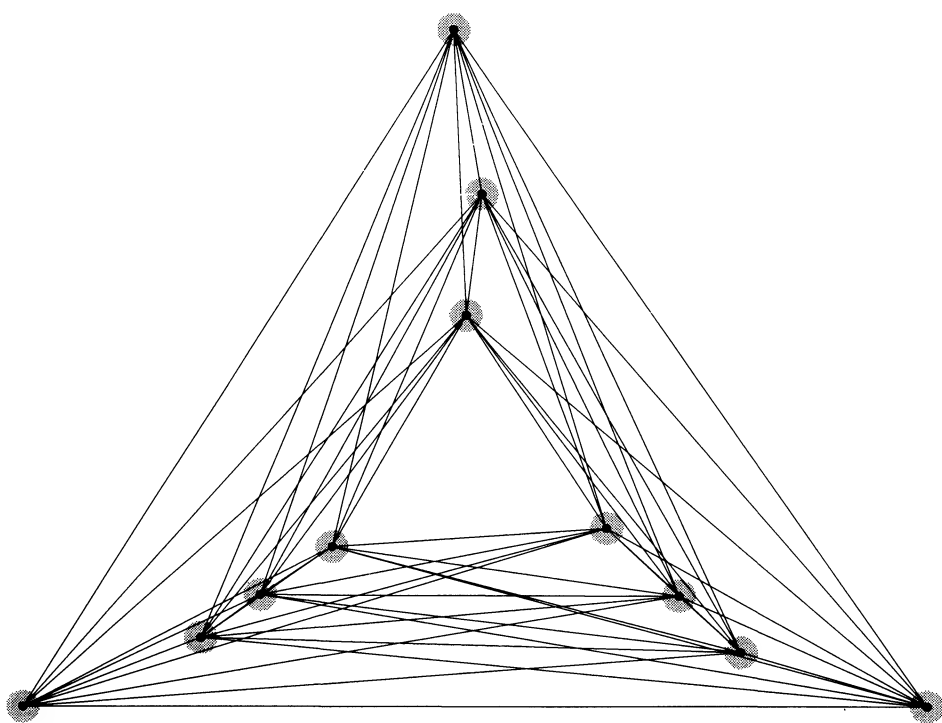
Steven C. Althoen, George Andrews, David F. Appleyard, Richard Askey, Duane W. Bailey, Tom Banchoff, James E. Baumgartner, James Berger, Paul Binding, David Bressoud, John Brillhart, Erik Brisson, John Brothers, Robert B. Burckel, David M. Burton, Gulbank D. Chakerian, Jingsen Chen, David A. Cox, John W. Dawson Jr., Manfred Denker, Robert Devaney, Persi Diaconis, Robert P. Dobrow, Todd A. Drumm, Bradley Efron, James Fill, Harley Flanders, Michael E. Gage, Theodore W. Gamelin, Ira M. Gessel, Andrew Glass, Andrew J. Granville, Branko Grünbaum, William Gustafson, Roger Horn, Norman W. Johnson, John W. Kenelly, Victor Klee, Frank Thomson Leighton, Jonathan W. Lewin, James G. Mauldon, Peter McMullen, William J. Mitchell, Michael Nauenberg, Alec Norton, Joseph O'Rourke, Ellen Parker, Robin A. Pemantle, Vera S. Pless, Gregory T. Quenell, Elmer Rees, Bruce Reznick, Gerhard Ringel, Joseph J. Rotman, Walter Rudin, David J. Rusin, Doris W. Schattschneider, Ron Sedgewick, Marjorie Senechal, Daniel Shanks, Joseph H. Silverman, David Allen Singer, Anita E. Solow, Lawrence E. Somer, Linda R. Sons, Norton Starr, Gilbert Strang, Craig Tracy, Alfred van der Poorten, Charles L. Vanden Eynden, Ellen Walker Lowenfield, Bruce Weide, Joel L. Weiner, Derick Wood, Lily Yen, Lawrence A. Zalcman, Doron Zielberger.

We give a special thanks to the Notes consultants: John Akeroyd, Mark Arnold, Dan Luecking, Mihalis Maliakas, Serge Tabachnikov.

The American Mathematical Monthly



Volume 101, Number 10 / DECEMBER 1994



AN OFFICIAL PUBLICATION OF THE MATHEMATICAL ASSOCIATION OF AMERICA

NOTICE TO AUTHORS

The *Monthly* publishes articles, notes, and other features about mathematics and the profession. The readership of the *Monthly* is intended to include everybody who is mathematically inclined, including of course professional mathematicians and students of mathematics at all collegiate levels. While no single article or feature is likely to appeal to everyone, material should interest and be accessible to a large number of readers. This is the most important criterion for acceptance.

Articles may be expositions of old results or presentations of new ones. They may concern all of mathematics or one small area, a broad development or a single application, historical reminiscences or one important event. While some articles may contain the author's new research, the novelty of material and generality of the results is far less important than the clarity of exposition and general interest. Discussing one illuminating case of a well known result is far better than providing all the details of an obscure but new proposition. Articles in the *Monthly* are supposed to inform and to entertain; they are meant to be read rather than archived.

Notes are short and possibly informal articles. A note may concern a clever new proof of an old theorem, a novel way to present tired material, or a lively discussion of a philosophical (but still mathematical) issue. Also, any topic is suitable, so long as it is related to mathematics. Because a note is short, the first few sentences are the most important part: They should explain the purpose and invite the reader in. Photographs or diagrams often will attract the reader's attention.

All articles and notes should be sent to the editor:

JOHN EWING
Department of Mathematics
Indiana University
Bloomington, IN 47405

Please send 3 copies, typewritten on only one side of the paper. Illustrations should be carefully drawn on separate sheets of paper in black ink; the original should be without lettering and two copies should have appropriate captions and lettering indicated.

Proposed problems or solutions should be sent to:

RICHARD BUMBY,
P.O. Box 10971
New Brunswick, NJ 08906-0971.

Please send 2 copies of all material, typewritten if possible.

Letters to the Editor, both for publication and for private reading, should be sent to the Editor at the address given above. Comments, including criticisms, are welcome, as are all suggestions for making the *Monthly* a lively, entertaining, and informative journal.

EDITOR:

JOHN H. EWING

ASSOCIATE EDITORS:

PETER BORWEIN
RICHARD BUMBY
DENNIS DETURCK
UNDERWOOD DUDLEY
JOHN DUNCAN
JOAN FERRINI-MUNDY
JOSEPH GALLIAN
STEVEN GALOVICH
RICHARD GUY
DARRELL HAILE
PAUL HALMOS
JOAN HUTCHINSON

FRED KOCHMAN
CATHERINE MCGEACH
RICHARD NOWAKOWSKI
ARNOLD OSTEBEE
LEE RUBEL
ABE SHENITZER
LYNN STEEN
STAN WAGON
DOUGLAS WEST
HERBERT WILF
SANDY ZABELL
PAUL ZORN

EDITORIAL ASSISTANT:

MISTY CUMMINGS

STAFF ARTIST:

MIKE CAGLE

QUOTE MASTER:

MARK WOODARD

Reprint permission:

MARCIA P. SWARD, Executive Director

Advertising Correspondence:

Ms. ELAINE PEDREIRA, Advertising Manager

Subscription correspondence, change of address, and other inquiries:

Membership / Subscriptions Department

All at the address:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036.

Microfilm Editions: University Microfilms International,
Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT. Copyrighted by the Mathematical Association of America (Incorporated), 1994, including rights to this journal issue as a whole and, except where otherwise noted, rights to each individual contribution. General permission is granted to Institutional Members of the MAA for noncommercial reproduction in limited quantities of individual articles (in whole or in part) provided a complete reference is made to the source. Second class postage paid at Washington, DC, and additional mailing offices. **Postmaster:** Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.

The American Mathematical Monthly

Volume 101, Number 10 / DECEMBER 1994
(ISSN 0002-9890)



Contents

ARTICLES

- The Rectilinear Crossing Number of a Complete Graph and Sylvester's "Four Point Problem" of Geometric Probability / EDWARD R. SCHEINERMAN
and HERBERT S. WILF 939
- String Matching for the Novice / ORA E. PERCUS and JEROME K. PERCUS 944
- Bernoulli Trials and Number Theory / DON RAWLINGS 948
- What Is the Shape of a Mylar Balloon? / WILLIAM H. PAULSEN 953
- Euler's Theorem for Polyhedra: A Topologist and Geometer Respond / PETER HILTON
and JEAN PEDERSEN 959
- The Role of Paradoxes in the Evolution of Mathematics / I. KLEINER
and N. MOVSHOVITZ-HADAR 963
- Regions in the Complex Plane Containing the Eigenvalues of a Matrix /
RICHARD A. BRUALDI and STEPHEN MELLENDORF 975
- Characterization of Solvable Quintics $x^5 + ax + b$ / BLAIR K. SPEARMAN
and KENNETH S. WILLIAMS 986
- A Halmos Problem and a Related Problem / JOHN B. COSGRAVE 993
-

FEATURES

COMMENTS 938

NOTES

- A "Popular" Class Number Formula / KURT GIRSTMAIR 997
- Variations on Wolstenholme's Theorem / EMRE ALKAN 1001
- The Second-Partials Test for Local Extrema of $f(x, y)$ /
LEONARD GILLMAN 1004

UNSOLVED PROBLEMS

- Mousetrap / RICHARD K. GUY
and RICHARD J. NOWAKOWSKI 1007

THE AUTHORS 1011

PROBLEMS AND SOLUTIONS 1013

REVIEWS

- Linear Programs and Related Problems*, by Evar D. Nering
and Albert W. Tucker / STEPHEN B. MAURER 1022

TELEGRAPHIC REVIEWS 1027

INDEX TO AMERICAN MATHEMATICAL MONTHLY VOLUME 101 1033

COMMENTS

Assessment is sweeping the country these days. Almost every section of the MAA has discussions of assessment. Accreditation agencies mandate that universities implement it. Deans direct departments to formulate plans. Chairs form committees to study the problem. And faculty discover that they've been doing assessment for years — they just haven't been doing it right.

It's hard to be against something so obviously sensible. Assessment means setting goals for programs and courses, and then measuring how effectively the goals are met. What knowledge and skills should be emphasized in undergraduate education? Are students learning what we think we are teaching? Those are pretty sensible questions, and they ought to be asked. I thought they were.

At a recent meeting, one person began his presentation by saying that the terminology and ideas of assessment may seem unfamiliar to mathematicians, but these were all ideas that had been in education schools for years. I wasn't comforted. Another speaker displayed a chart that showed *current* assessment practices in his college. All departments were listed vertically, and across the page were various kinds of assessment activities. Most departments had a few things checked. *One* had twice as many as the others — it was Education. They seem to be more advanced than the rest of us.

In much of the midwest, assessment is driven by the North Central Association. Universities and colleges must put in place elaborate assessment plans or risk their accreditation. They hire outside assessment experts who run workshops and train local assessment experts. I wonder where these people first learned to be experts?

Part of my uneasiness about assessment is the lack of precision. Here are some phrases from the *Handbook of Assessment* that my university provides to all departments. "Assessment focuses on student learning outcomes." Learning outcomes? "Assessment is research and inquiry into the improvement of teaching and learning." What's it mean to inquire into improvement? "Assessment is a process in which goals and learning objectives of a program or course are identified and data are collected from multiple sources to document student, teacher, or program achievement of those goals and objectives." It seems strange that the same handbook advises us when creating *instruments* to "choose words that have precise meanings" and to "avoid complex or awkward syntax." What *is* an instrument anyway?

A basic goal of assessment seems to be measuring the value added to students by attending our classes. This emphasizes the manufacturing approach to education. I wonder how one would assess a trip to the museum or a child's visit to the playground or a night at the symphony. What are the goals? Were the objectives met? I'd like to think that attending a university is closer to a trip to a museum or a playground than a trip along an assembly line.

One of the benefits of assessment is the "added responsibility teachers feel for student learning." Of course, teachers *ought* to be responsible. But I wish I knew of some technique to make *students* feel more responsible for student learning.

John Ewing

The Rectilinear Crossing Number of a Complete Graph and Sylvester's "Four Point Problem" of Geometric Probability

Edward R. Scheinerman and Herbert S. Wilf

The chance of ... the quadrilateral formed by joining four points, taken arbitrarily within any assigned boundary, constituting a reentrant or convex quadrilateral, will serve as types of the class of questions in view.

—J. J. Sylvester [11]

We prove that two fundamental constants of the geometry of the plane are equal.

First, if R is an open set in the plane with finite Lebesgue measure, let $q(R)$ denote the probability that if four points are chosen independently uniformly at random in R , then their convex hull is a quadrilateral. Let q_* be the infimum of $q(R)$ over all such sets R .

Second, let $\bar{\nu}(K_n)$ denote the rectilinear crossing number of the complete graph on n vertices, i.e., the minimum number of intersections in any drawing of K_n in the plane that has straight-line-segment edges. It is well known that $\bar{\nu}(K_n)/\binom{n}{4}$ increases steadily to some limit $\bar{\nu}^*$ as $n \rightarrow \infty$.

Our main result is that $q_* = \bar{\nu}^*$.

FOUR RANDOM POINTS. Let R be an open set in the plane with finite area. As such, we can consider R to be a sample space from which we select points independently uniformly at random (i.u.a.r.). Choose four points from R i.u.a.r.. Then with probability 1, no three of the points are collinear, so the convex hull of the four points is either a triangle (one point in the convex hull of the other three) or a quadrilateral. J. J. Sylvester [11] asked, what is the probability that the points determine a convex quadrilateral? We denote this probability by $q(R)$.

How large and how small can $q(R)$ be? When R is restricted to being a convex set we have the following result (see Blaschke [1, 2] and also [7]).

Theorem 1. *Let R be an open, convex subset of the plane, of finite area.*

Then

$$\frac{2}{3} \leq q(R) \leq 1 - \frac{35}{12\pi^2} \approx 0.704.$$

Further, both inequalities are sharp. The lower bound is attained by the interior of a triangle (any triangle), and the upper bound by the interior of an ellipse (any ellipse).

□

This theorem, however, does not fully address the issue of the extreme values of $q(R)$ because it considers only *convex* regions R . It is easy to see that if we relax

the convexity requirement, then the supremum of $q(R)$ is 1; let R be a very thin open annulus and observe that we can make $q(R)$ arbitrarily close to 1. Thus it remains to consider the infimum of $q(R)$; let $q_* = \inf q(R)$ where the infimum is over all open sets R with finite area.

We show below that q_* is positive and strictly less than $2/3$ (the lowest possible result for convex R). We show further that q_* is closely related to the rectilinear crossing number of complete graphs.

RECTILINEAR CROSSING NUMBER OF A GRAPH. Let G be a graph which we wish to draw in the plane. If G is planar, then we can find an embedding in which the edges do not cross. A result of Fáry [4] shows that we can choose this embedding so that the edges are noncrossing straight line segments.

Let $\bar{\nu}(G)$ denote the minimum number of crossings in a straight line drawing of G in the plane; the parameter $\bar{\nu}(G)$ is known as the *rectilinear crossing number* of G . (For background on the rectilinear crossing number, see [6], [7] or [12].)

An important open problem in the study of graph embeddings is to determine the rectilinear crossing number of the complete graph K_n . For $n = 5, 6, 7, 8, 9$ the values are known (see [12]) and they are 1, 3, 9, 19, 36 respectively. For $n = 10$ it is known [10] that $61 \leq \bar{\nu}(K_{10}) \leq 62$.

If we place the n vertices of K_n on a circle, then the number of crossings is exactly $\binom{n}{4}$; certainly we can do better, but $\bar{\nu}(K_n)$ is on the order of n^4 as we now explain.

Theorem 2. *There exists a constant $\bar{\nu}^*$ such that $0 < \bar{\nu}^* < \infty$ and*

$$\bar{\nu}^* = \lim_{n \rightarrow \infty} \frac{\bar{\nu}(K_n)}{\binom{n}{4}} = \sup_n \frac{\bar{\nu}(K_n)}{\binom{n}{4}}.$$

(This is well-known folklore, but for completeness we show the proof here.)

Proof: Let $m < n$ and consider a straight line embedding of K_n in the plane with the minimum number of crossings, $\bar{\nu}(K_n)$. For each m element subset A of $V(K_n)$, let $c(A)$ denote the number of crossings in this embedding in which the endpoints of the crossing edges are all in A . If we sum $c(A)$ over all m -subsets of $V(K)$, we count each possible crossing exactly $\binom{n-4}{m-4}$ times. Thus

$$\bar{\nu}(K_n) = \sum_{|A|=m} c(A) \bigg/ \binom{n-4}{m-4}.$$

Now clearly $c(A) \geq \bar{\nu}(K_m)$, so it follows that

$$\bar{\nu}(K_n) \geq \frac{\binom{n}{m}}{\binom{n-4}{m-4}} \bar{\nu}(K_m)$$

which we can rearrange to read

$$\frac{\bar{\nu}(K_n)}{\binom{n}{4}} \geq \frac{\bar{\nu}(K_m)}{\binom{m}{4}}.$$

Thus $\bar{\nu}(K_n)/\binom{n}{4}$ is a nondecreasing function of n which is bounded above by 1 and below by $\bar{\nu}(K_5)/\binom{5}{4} = 1/5$. \square

Since $\bar{\nu}(K_{10}) \geq 61$, we see that $\bar{\nu}^* \geq 61/210 \approx 0.29$.

Singer [10] proves the following upper bound on the rectilinear crossing number of K_n when n is a power of 3:

$$\bar{\nu}(K_n) \leq \frac{1}{312}(5n^4 - 39n^3 + 91n^2 - 57n).$$

Thus $\bar{\nu}^* \leq \frac{5}{312} \times 24 = \frac{5}{13} \approx 0.3846$.

(Jensen [6] gives a rectilinear embedding of K_n with $\frac{7}{432}n^4 + O(n^3)$ crossings, yielding an upper bound of 0.3888... on $\bar{\nu}^*$.)

MAIN RESULT. Our main result is a simple relation between q_* , the smallest probability of choosing a quadrilateral, and $\bar{\nu}^*$, the limit of $\bar{\nu}(K_n)/\binom{n}{4}$.

Theorem 3. *With the preceding notation, $q_* = \bar{\nu}^*$.*

Proof: Let R be any open set in the plane with finite area. Choose n points p_1, p_2, \dots, p_n i.u.a.r. from R and let those n points be the vertices of a straight line drawing of K_n . Let c be the number of crossings in this drawing. Now c is a random variable whose value is always at least $\bar{\nu}(K_n)$. Further, let

$$X = \sum_{\{a, b, c, d\}} \mathbf{1}\{p_a, p_b, p_c, p_d \text{ form a quadrilateral}\}$$

where the sum is over all 4 element subsets of $\{1, \dots, n\}$ and $\mathbf{1}\{\dots\}$ is a 0, 1 indicator random variable whose value is 1 just when the convex hull of the four points p_a, p_b, p_c, p_d is a quadrilateral. Since the optimum drawing cannot have more crossings than the average, we get, by taking expectations,

$$\bar{\nu}(K_n) \leq E(X) = \binom{n}{4} q(R)$$

for all n . Dividing by $\binom{n}{4}$ and letting $n \rightarrow \infty$, we have $\bar{\nu}^* \leq q(R)$ for all R . Thus $\bar{\nu}^* \leq q_*$.

For the opposite inequality, consider a straight line embedding of K_n with the minimum number, $\bar{\nu}(K_n)$, of crossings. Let R_ϵ be the (disconnected) open set formed by placing a small open disc of radius ϵ centered at each vertex of the embedding. See the cover of this issue. Here ϵ is chosen small enough so that for every choice of n points, one in each disc, if we connect all pairs of them by straight line segments then the number of crossings is always equal to $\bar{\nu}(K_n)$, i.e., all such embeddings are optimal. Clearly such an ϵ exists.

We now consider the following question: choose four distinct discs of R_ϵ , and then choose i.u.a.r. a point from each of them. What is the probability q that the resulting quadrilateral is convex?

On the one hand, q is the number of convex quadrilaterals in the original embedding divided by $\binom{n}{4}$. But the former are in 1-1 correspondence with edge crossings, so there are exactly $\bar{\nu}(K_n)$ of them, and we have $q = \bar{\nu}(K_n)/\binom{n}{4}$.

On the other hand, $q(R_\epsilon)$ is the probability that four points chosen i.u.a.r. in R_ϵ will form a convex quadrilateral. But four points so chosen will lie in four distinct

discs of R_ϵ with probability $1 - O(1/n)$. Hence

$$q = q(R_\epsilon) + O(1/n) \geq q_* + O(1/n).$$

Combining these two facts about q we obtain

$$q = \bar{\nu}(K_n) / \binom{n}{4} \geq q_* + O(1/n).$$

If we now let $n \rightarrow \infty$, $\bar{\nu}^* \geq q_*$ follows, and the proof is complete. \square

Thus to summarize our principal results, we have

$$0.29 \approx \frac{61}{210} \leq q_* = \bar{\nu}^* \leq \frac{5}{13} \approx 0.385.$$

SOME COMMENTS

1. In a spirit similar to ours, Moon [8] applies random methods in bounding the ordinary crossing number of K_n . He places n points i.u.a.r. on a sphere and joins them pairwise by arcs of great circles. This gives an upper bound of $\frac{3}{8}\binom{n}{4}$ for the crossing number of K_n . It is not clear how to project Moon's embedding into the plane and have the edges become line segments.
2. Our methods can be applied to arbitrary graphs G . Let M denote the number of pairs of edges in G which span four distinct vertices. Then $\bar{\nu}(G) \leq \bar{\nu}^* M/3$; simply place the vertices of G i.u.a.r. in a region R and compute the expected number of crossings.
3. Given a subset R of the plane together with a probability measure μ defined on R , define $q(R, \mu)$ to be the probability that four points chosen i.u.a.r. with respect to μ form a convex quadrilateral. Without further restrictions, we see that the infimum of $q(R, \mu)$ is 0; let R be a unit line segment together with Lebesgue measure—no four points can form a quadrilateral.

If we restrict ourselves to those (R, μ) for which the probability that three points selected i.u.a.r. are collinear is 0, then the infimum of $q(R, \mu)$ remains the same, namely q_* .

If we understand Sylvester's problem (see quote above) to mean the interior of a Jordan curve, our results still don't change; in the proof of Theorem 3 we can join the small open disks by even smaller tendrils so the domain is the interior of a simple closed curve.

4. Let R be a bounded open set and embed the vertices of K_n at n points selected i.u.a.r. in R . We have seen that the expected number of crossings in this embedding is $q(R)\binom{n}{4}$. However, one might harbor hopes of doing better on occasion. It would seem natural to generate many embeddings of K_n in, say, the interior of a square or a disk, and count the number of crossings in hopes of finding a good embedding. Regrettably, this is not at all likely, as we now explain.

The number of crossings can be written

$$X = \sum_{\{a, b, c, d\}} \mathbf{1}\{p_a, p_b, p_c, p_d \text{ form a quadrilateral}\}$$

where the sum is over all 4-element subsets of $\{1, \dots, n\}$ and the p_i 's are chosen i.u.a.r. in R . Thus X is an example of a *U-statistic*; see [5] or [9] for an extensive discussion. Using "deviations" results (see [9] §5.6) one can

show that the probability that the number of crossings is “significantly” less than the expectation is extremely small.

REFERENCES

1. W. Blaschke, Über affine Geometrie XI: Lösung des “Vierpunktproblems” von Sylvester aus der Theorie der geometrischen Wahrscheinlichkeiten, *Leipziger Berichte* 69 (1917) 436–453.
2. W. Blaschke, *Vorlesungen über Differentialgeometrie II: Affine Differentialgeometrie*, Springer, Berlin (1923).
3. P. Erdős, and R. K. Guy, Crossing number problems, *This Monthly* 80 (1973) 52–57.
4. I. Fáry, On straight line representations of planar graphs, *Acta Sci. Math. (Szeged)* 11 (1948) 229–233.
5. W. Hoeffding, The strong law of large numbers for U -statistics, Univ. North Carolina Institute of Statistics Mimeo Series, no. 302 (1961).
6. H. F. Jensen, An upper bound for the rectilinear crossing number of the complete graph, *J. Comb. Theory (B)* 10 (1971) 212–216.
7. Victor Klee, What is the expected volume of a simplex whose vertices are chosen at random from a given convex body?, *This Monthly* 76 (1969) 286–288.
8. J. W. Moon, On the distribution of crossings in random complete graphs, *SIAM J.* 13 (1965) 506–510.
9. Robert J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley (1980).
10. David Singer, The rectilinear crossing number of certain graphs, manuscript (1971).
11. J. J. Sylvester, On a special class of questions on the theory of probabilities, *Birmingham British Association Report* (1865) 8.
12. Arthur T. White, and Lowell W. Beineke, Topological graph theory, in *Selected Topics in Graph Theory*, Beineke & Wilson, eds., Academic (1978) 15–49.

Department of Mathematical Sciences
The Johns Hopkins University
Baltimore, MD 21218-2689
ers@cs.jhu.edu

Department of Mathematics
University of Pennsylvania
Philadelphia, PA 19104-6395
wilf@central.cis.upenn.edu

Biographical history, as taught in our public schools, is still largely a history of boneheads: ridiculous kings and queens, paranoid political leaders, compulsive voyagers, ignorant generals—the flotsam and jetsam of historical currents. The men who radically altered history, the great scientists and mathematicians, are seldom mentioned, if at all.

—*Martin Gardner*

George F. Simmons, *Calculus Gems*.
 New York: McGraw Hill, Inc., 1992, p. 1

String Matching for the Novice

Ora E. Percus and Jerome K. Percus

A recurrent theme to gladden the hearts of jaded mathematicians is the frequent rediscovery, for the purposes of today's technology, of elegant mathematical results from years, decades, or even centuries ago. The development of quantum mechanics via the "matrix mechanics" of Born and Heisenberg in (anecdotal?) ignorance of the very well developed theory of matrices, is a widely quoted example. But technology today often has the prefix "bio" attached to it, and so it is only fitting to find that statistical questions arising in the human genome project were considered, and solved in practical form, centuries ago. It is the purpose of this note to go one step further and show that even the naive mathematical level that must have been the norm in bygone days is sufficient for practical answers to these same questions.

The questions we have in mind are all versions of "the matching problem". A typical context is this. (See e.g. ref. 5.) A large linear molecular chain occurs as a string of l molecular units (amino acids for proteins, bases for DNA, ...); another molecule of length l when aligned with the first is found to have a subsequence of r units in common with it. What is the a priori probability of this event, to be used for example in a statistical test of the hypothesis that this was a purely random occurrence, and hence not a significant indication of some functional or evolutionary relationship between the two molecules? Now the "randomness" assumption has to be made more precise. For this purpose, we suppose that the relative frequency f_j of the j th type of unit, $j = 1, \dots, n$ (proteins: $n = 20$, DNA: $n = 4, \dots$) is known, and that this is the frequency at which type j is to be found at a given location or site in a molecular string, independently of the identity of the units at other sites. If this is the case, the probability of units from two molecules matching at a given site will be

$$p = \sum_{j=1}^n f_j^2, \quad (1)$$

and the probability of their not matching, $q = 1 - p$.

PROTOTYPE. Let us return to the question at hand, now framed as: what is the probability that, under random selection, the two strings of units will match at least at r sites in a row? This was studied from a generating function viewpoint by de Moivre [1] back in the 1700's, and an explicit highly computable series result is to be found, e.g. in Uspensky's standard text [6]. The conceptual difficulty is that e.g. failure to match a given r -site subsequence biases the matching probability of another r -subsequence with overlapping sites. In other words, two successful matching events A and B for different subsequences will *not* exactly satisfy

$$P(A \cup B) = P(A) + P(B) \quad (2)$$

because they are not mutually exclusive. There are ingenious ways of bounding the effect of such potential overlaps on the resulting probability, but here we simply note that if the r -matching probability is very small to start with—in accord with many practical situations—then we anticipate that the overlap correction will be extremely small and can be neglected. In the present case, the implications are immediate: the probability of at least r matches starting at the left end (designated site 1) of the pair of molecules is clearly p^r ; the probability of an r -match starting at site $s = 2, \dots, l + 1 - r$ is that of site $s - 1$ having a mismatch, followed by r matches, i.e. qp^r , and there are $l - r$ such values of s . Hence the total probability of at least one r -match under the assumption of additive probabilities is given by

$$\begin{aligned} P_{\geq r}^l &= p^r + (l - r)qp^r \\ &= (1 + (l - r)q)p^r, \quad \text{for } r \geq 1. \end{aligned} \quad (3)$$

We can just as easily be a little more precise. Suppose that we ask for P_r^l , the probability that the *longest* match has length r , i.e. that there are length r matches and they do not extend to length $r + 1$. This means that unless we are at either end, a mismatch followed by r matches followed by a mismatch is required, a probability of $qp^r q$ for $s = 2, \dots, l - r$; starting at $s = 1$, one needs only $p^r q$, and finishing at $s = l$, only qp^r . Thus, $P_r^l = p^r q + (l - r - 1)qp^r q + qp^r$, or

$$P_r^l = \begin{cases} (2 + (l - 1 - r)q)p^r q & \text{for } 1 \leq r < l \\ p^l & \text{for } r = l. \end{cases} \quad (4)$$

Of course, (3) and (4) are related. The probability of the longest match having length r is that of at least r in a row minus that of at least $r + 1$ in a row, and indeed $(1 + (l - r)q)p^r - (1 + (l - r - 1)q)p^{r+1} = (2 + (l - 1 - r)q)p^r q$ is an identity for $r < l$, and holds trivially for $r = l$.

NEXT STAGE. This was a warm-up. Since the long molecules are built up of bits and pieces which get arranged and rearranged, there is no reason to expect a common functional subsequence to be at the same relative set of locations in each molecule, and indeed no reason to expect the two potentially related molecules to have the same length. So let us suppose that the molecules are $l_2 \leq l_1$ units long, and that we ask for the probability of a match of at least r units in sequence of one with the same number of units in sequence some place in the other. The counting is most easily done by figuratively sliding one molecule along the other and regarding the common region of sites as a pair of equal length molecules to which we can apply (3). The length of this region will be l_2 any time the l_2 molecule can fit inside the l_1 molecule, i.e. $l_1 + 1 - l_2$ times, each contributing $(1 + (l_2 - r)q)p^r$ to the desired probability (again with the assumption of independence). Then, as the l_2 string moves partly outside the l_1 string at one end or the other, an overlap of $r \leq l < l_2$ will occur exactly twice, with a contribution of $(1 + (l - r)q)p^r$. The total probability is then $(l_1 + 1 - l_2)(1 + (l_2 - r)q)p^r + 2\sum_{l=r}^{l_2-1} (1 + (l - r)q)p^r$, which reduces without difficulty to

$$P_{\geq r}^{l_1, l_2} = [1 + (l_1 - r) + (l_2 - r) + (l_1 - r)(l_2 - r)q]p^r, \quad \text{for } 1 \leq r < l_2. \quad (5)$$

For example [4], two fragments of DNA, of lengths $l_1 = 154$, $l_2 = 103$ with nominal base matching probability $p = 1/4$ would have a match of at least nine bases length with probability $P_{\geq 9} = 0.040$, very close to the exact 0.039.

Of course, the corresponding result for the longest match being of length r can be found precisely as before, i.e. by computing $P_r^{l_1, l_2} = P_{\geq r}^{l_1, l_2} - P_{\geq r+1}^{l_1, l_2}$, and this

turns out to be

$$P_r^{l_1, l_2} = [2 + 2(l_1 - r + l_2 - r - 1)q + (l_1 - r - 1)(l_2 - r - 1)q^2] p^r, \quad \text{for } 1 \leq r < l_2 - 1. \quad (6)$$

RELAXED CRITERIA. It is well known that some substitutions can be tolerated without changing the functionality of biological subsequences. Suppose that the matching frame is of length r or greater, but that only m sites of a subsequence have to be identical, including of course the first and last sites that define the length r . Now the elementary matching probability, instead of being p^r , is determined by one match, $m - 2$ out of $r - 2$ matches, then another match, i.e. a probability of $p \binom{r-2}{m-2} p^{m-2} q^{r-m} p$. We conclude that the only effect of the relaxed criterion is the replacement

$$\text{for } m \text{ out of } r, \text{ replace: } p^r \rightarrow \binom{r-2}{m-2} p^m q^{r-m} \quad (7)$$

in (2), (3), (4), or (5). Presumably more relevant would be the criterion that at least m out of r match, leading at once to

$$\text{for at least } m \text{ out of } r, \text{ replace: } p^r \rightarrow \sum_{k=m}^r \binom{r-2}{k-2} p^k q^{r-k}. \quad (8)$$

Still more relevant might be a weighted probability in which the pattern of matches within an r -subsequence matters, so that e.g. one would bias against large gaps, but this is a matter of explicit relevant biochemistry and will not be discussed here.

ASSESSMENT. The assumption that all successful r -matches are nonoverlapping is clearly an approximation that holds only if the elementary events have sufficiently small probabilities that one is confident that the overlap probability is fully negligible. But how small is small? In this respect, a number of comparisons have been made between the results of the independence assumption and sophisticated bounding techniques or numerical simulations, arriving at the conclusion that “sufficiently small” can be quite large, to the extent that it is automatically satisfied by virtually any situation in which matching is unusual enough to warrant its use to contradict a hypothesis of randomness. So a naive approach is justified. But suppose one really wants to assess the possible effect of non-empty intersection of relevant events. How difficult is this? Not very, as a prototypical example shows.

Consider the basic matching problem as approximated by (3). If A_j is the event that a match at $\geq r$ contiguous sites starts at site j , $j = 1, 2, \dots, l + 1 - r$, then

$$P_{\geq r}^l = P(A_1 \cup A_2 \cup \dots \cup A_{l+1-r}). \quad (9)$$

One knows from Bonferroni's inequalities [2] that, quite generally,

$$\begin{aligned} \sum_{j=1}^{l+1-r} P(A_j) &\geq P\left(\bigcup_{j=1}^{l+1-r} A_j\right) \\ &\geq \sum_{j=1}^{l+1-r} P(A_j) - \sum_{1 \leq j < k \leq l+1-r} P(A_j \cap A_k). \end{aligned} \quad (10)$$

We have already used $P(A_1) = p^r$, $P(A_j) = qp^r$ for $j > 1$ in (3). Now if $j < k \leq j + r$, A_k overlaps or abuts A_j ; thus k cannot be the start of a match of exactly r ,

and $P(A_j \cap A_k) = 0$. On the other hand, if $k > j + r$, A_j and A_k are not mutually exclusive but independent, so that $P(A_j \cap A_k) = P(A_j)P(A_k)$. Thus the nonvanishing contributions come from: $j = 1$, $r + 1 < k \leq l + 1 - r$, a total of $(l - 2r)p^r qp^r$, and: $1 < j \leq l - 2r$, $r + j < k \leq l + 1 - r$, a total of $\sum_{j=2}^{l-2r} (l + 1 - 2r - j)(qp^r)^2 = \frac{1}{2}(l - 2r - 1)(l - 2r)(qp^r)^2$. It follows from (3) and (10) that

$$(1 + (l - r)q)p^r \geq P_{\geq r}^l \geq (1 + (l - r)q)p^r - (l - 2r)(1 + \frac{1}{2}q(l - 2r - 1))qp^{2r}. \quad (11)$$

Interestingly, the terms in (11) are precisely the first two in Uspensky's expansion (and the remaining terms can be obtained as well using inclusion-exclusion). A quoted example [3] is in the context of the tossing of a fair coin, $p = 1/2$. For 2059 tosses, the probability of a run of at least 13 is given exactly by .1176. According to (11), one finds $.1249 \geq P_{\geq 13}^{2059} \geq .1172$, illustrating as well the rapid convergence of the alternating series that equation (10) is developing.

REMARKS. The above argument suggests an alternative strategy, which is to start by assuming that successful r -matches are independent events. Then (9), in the form $P_{\geq r}^l = 1 - P(\bar{A}_1 \cap \bar{A}_2 \dots)$ where \bar{A} is the complement of A , yields

$$P_{\geq r}^l = 1 - \prod_i (1 - P(A_i)), \quad (12)$$

which takes on a Poisson distribution form for small p^r , coincides with the lower bound of (11) when $r \ll l$, and can be sequentially corrected for correlations. Establishing bounds in this fashion is not trivial. Of course, bounding correction terms can be derived for other cases, just as they were for (11), with similar qualitative results: the correction ΔP to the probability P scales as the square of the probability (here $P \sim lqp^r$, $\Delta P \sim \frac{1}{2}(lqp^r)^2$), validating our initial presumption as to the effect of additivity of probabilities. However, since we have for example omitted correlations among sites on the *same* chain, which are known to exist, it is not clear that we are justified in going beyond the mathematics of the early days of probability, as illustrated by (3), (4), (5), and (6) which give us the upper bounds that tell us whether an observation is likely to represent nonrandomness. And this is the message which was to be delivered.

REFERENCES

1. A. De Moivre, *Doctrine of Chance*, 3rd ed. (1756).
2. W. Feller, *Probability Theory and its Applications*, Vol. 1, John Wiley & Sons, Inc., 2nd Edition, p. 100 (1950).
3. L. Goldstein, Poisson approximation and DNA sequence matching, *Commun. Stat. Theor. Meth.* 19(11), 4167-4179 (1990).
4. L. Goldstein, and M. S. Waterman, Poisson, compound Poisson, and process approximations, *Bull. Math. Bio.* 54, 785-812 (1992).
5. R. F. Mott, T. B. L. Kirkwood, and R. N. Curnow, An accurate approximation to the distribution of the length of the longest matching word between two random DNA sequences, *Bull. Math. Bio.* 52, 773-784 (1990).
6. J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill, NY (1937).

Courant Institute of Mathematical Sciences
New York University
251 Mercer Street
New York, NY 10012

Bernoulli Trials and Number Theory

Don Rawlings

1. INTRODUCTION. A coin may be used to generate a natural number by simply tossing it until heads lands up: If the heads occurs on the n th toss, then identify the sequence of tosses with n . For instance, the Bernoulli sequence TTTTTH may be viewed as generating the natural number 6. The probability $P_q(n)$ that a given number n is the outcome may be readily determined. If the coin lands tails up with probability $q < 1$, then $P_q(n) = q^{n-1}(1 - q)$ since the trials are independent.

Of course, the set of natural numbers Z^+ along with the measure P_q really amounts to the geometrical distribution with success probability $(1 - q)$. However, the perspective from the measure space (Z^+, P_q) is provocative; it serves as a catalyst for the consideration of questions with a number theoretic flavor in the context of a geometrical distribution.

Two examples illustrating the possibilities afforded by this perspective are presented herein. In the first, a certain divisibility property considered on Bernoulli generated s -tuples of natural numbers leads to a natural “ q -analog” of Euler’s product formula for the Riemann zeta function. The second example concerns the probability of a sequence n_1, n_2, \dots, n_s of Bernoulli generated numbers having no “fixed indices.” As $s \rightarrow \infty$, the likelihood of having no “fixed indices” will be shown to approach the reciprocal of a “ q -analog” of the real number e .

2. THE NOTION OF q -ANALOG. An object $\mathcal{O}(q)$ is said to be a q -analog of \mathcal{O} if $\mathcal{O}(q)$ reduces to \mathcal{O} when $q = 1$. As an illustration, for $n \in Z^+$ the polynomials defined by

$$[n] = 1 + q + q^2 + \cdots + q^{n-1} \quad \text{and} \quad [n]! = [1][2] \cdots [n]$$

are q -analogs respectively of n and its factorial. These q -analogs are extended to include 0 by adopting the conventions that $[0] = 0$ and $[0]! = 1$. The polynomial $[n]$ is just the n th partial sum of a geometric series in q .

An important example of a q -analog for the discourse at hand is the continuous extension on congruence classes of P_q to include the case $q = 1$. For $m \in Z^+$, let $\bar{1}, \bar{2}, \dots, \bar{m}$ denote the set of congruence classes modulo m . Since $[m] = (1 - q^m)/(1 - q)$, it follows for $0 \leq q < 1$ that the probability of a Bernoulli generated number n belonging to the class \bar{r} is

$$\begin{aligned} P_q\{n \in \bar{r}\} &= \sum_{k=0}^{\infty} q^{r-1+km}(1 - q) \\ &= (1 - q)q^{r-1} \sum_{k=0}^{\infty} (q^m)^k = \frac{(1 - q)q^{r-1}}{1 - q^m} = \frac{q^{r-1}}{[m]}. \end{aligned}$$

Since $q^{r-1}/[m]$ is continuous from the left at $q = 1$, P_q may be continuously extended on congruence classes to $q = 1$ by defining $P_1\{n \in \bar{r}\} = 1/m$ for $1 \leq r \leq m$. Thus, the extended P_q is a q -analog of the equiprobable measure.

Actually, this extension makes some sense in terms of the Bernoulli scheme: If $q = 1$, then the coin almost surely never lands heads up. So, all numbers in Z^+ are equally likely (or, more aptly speaking, equally unlikely). For later reference, the remarks of this paragraph and the last are recorded in the following lemma.

Lemma 1. For $0 \leq q \leq 1$, the probability of a Bernoulli generated number n belonging to the congruence class \bar{r} modulo m is $P_q\{n \in \bar{r}\} = q^{r-1}/[m]$.

3. A q -ANALOG OF THE RIEMANN ZETA FUNCTION. For $n \in Z^+$, suppose that $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$ where p_1, p_2, \dots, p_k are distinct primes and $\alpha_1, \alpha_2, \dots, \alpha_k$ are positive integers. The q -canonical factorization of n is then defined to be

$$\{n\} = [p_1]^{\alpha_1} [p_2]^{\alpha_2} \cdots [p_k]^{\alpha_k}.$$

Furthermore, let $\nu(n) = \alpha_1(p_1 - 1) + \alpha_2(p_2 - 1) + \cdots + \alpha_k(p_k - 1)$. By convention, set $\{1\} = 1$ and $\nu(1) = 0$.

For $s > 1$ and $0 \leq q \leq 1$, a q -analog of the Riemann zeta function may be defined by

$$\zeta_q(s) = \sum_{n=1}^{\infty} \frac{q^{s\nu(n)}}{\{n\}^s}.$$

Note that $\zeta_q(s)$ does indeed reduce to the Riemann zeta function $\zeta(s)$ when $q = 1$. As might be expected, $\zeta_q(s)$ has some properties that are analogous to those of $\zeta(s)$. The relevant ones for the purposes of the next section are tabulated in the next two theorems:

Theorem 1. For $s > 1$ and $0 \leq q \leq 1$, $\zeta_q(s)$ is convergent.

Theorem 2. For $s > 1$ and $0 \leq q \leq 1$, $\zeta_q(s) = \prod_p (1 - q^{s(p-1)}/[p]^s)^{-1}$ where the product is over all primes p .

Theorem 2 is a q -analog of Euler's classical product formula for the Riemann zeta function. Rather than getting momentarily mired down in analytic details, the proofs of Theorems 1 and 2 are postponed until sections 6 and 7. Let's now turn our attention to the context that served to motivate Theorem 2.

4. p -DIVISIBILITY AND BERNOULLI GENERATED s -TUPLES. For p prime, an s -tuple (n_1, n_2, \dots, n_s) of natural numbers is said to be p -divisible if p divides each n_j . By considering the notion of p -divisibility in the context of Bernoulli trials, a probabilistic application of $1/\zeta_q(s)$ arises.

To see how this comes about, consider repeating the Bernoulli scheme of the introduction s times for each prime p . The result will be a sequence of s -tuples, each of the form $(n_1, n_2, \dots, n_s)_p$. Since the components of such an s -tuple are determined independently, Lemma 1 implies that the probability of $(n_1, n_2, \dots, n_s)_p$ not being p -divisible is

$$1 - \prod_{j=1}^s P_q\{n_j \in \bar{p}\} = 1 - \prod_{j=1}^s \frac{q^{p-1}}{[p]} = 1 - q^{s(p-1)}/[p]^s.$$

Furthermore, provided that $s > 1$, the probability of $(n_1, n_2, \dots, n_s)_p$ not being

p -divisible for all primes p is given by

$$\prod_p (1 - q^{s(p-1)}/[p]^s).$$

By Theorem 2, this infinite product converges to $1/\zeta_q(s)$. These considerations are summarized in the following theorem.

Theorem 3. Suppose that $s > 1$ and that the Bernoulli scheme of the introduction is repeated s times for each prime p so as to generate a sequence of s -tuples, each of the form $(n_1, n_2, \dots, n_s)_p$. The probability of no $(n_1, n_2, \dots, n_s)_p$ being p -divisible is $1/\zeta_q(s)$.

5. SEQUENCES WITH NO FIXED INDICES. Before addressing the subject of this section, it is expedient to first dispose of a few technical details. In [4], a q -analog of the real number e is defined as the series

$$[e] = \sum_{n=0}^{\infty} \frac{1}{[n]}!$$

which converges for $0 < q \leq 1$. If $q = 0$, then $[e] = \infty$. As demonstrated in [4], this q -analog of e satisfies the following limit formula:

Theorem 4. If $0 \leq q \leq 1$, then

$$[e] = \lim_{n \rightarrow \infty} \prod_{j=1}^n \left(1 - \frac{q^{j-1}}{[n]}\right)^{-1}$$

This is just a q -analog of the well-known identity $e = \lim_{n \rightarrow \infty} (1 - 1/n)^{-n}$. The stage is now set.

A sequence n_1, n_2, \dots, n_s of natural numbers is said to have *no fixed indices modulo s* if $n_j \notin \bar{j}$ for $1 \leq j \leq s$. Although not standard fare for number theory, this notion leads to an identity involving the reciprocal of $[e]$.

By making use of independence and of Lemma 1, the probability that a Bernoulli generated sequence n_1, n_2, \dots, n_s has no fixed indices modulo s is

$$\prod_{j=1}^s P_q\{n_j \notin \bar{j}\} = \prod_{j=1}^s \left(1 - P_q\{n_j \in \bar{j}\}\right) = \prod_{j=1}^s \left(1 - \frac{q^{j-1}}{[s]}\right).$$

In view of Theorem 4, letting $s \rightarrow \infty$ yields

Theorem 5. The probability that a Bernoulli generated sequence n_1, n_2, \dots, n_s has no fixed indices approaches $1/[e]$ as $s \rightarrow \infty$.

Theorem 5 was inspired by and is equivalent to a result in [4] concerning “ q -random mappings” with no fixed points.

6. A PROOF OF THEOREM 1. Theorem 1 is clearly true for the extreme values of q : On the one hand, $\zeta_0(s) = 1$. On the other, $\zeta_1(s)$ is just the Riemann zeta function which is known to converge for $s > 1$.

So let's now restrict our attention to the case $0 < q < 1$. To get at this case, there are a few facts that are best isolated. They are presented in Lemma 2.

Lemma 2. If $0 < q < 1$ and $m > 1$, then

$$\begin{aligned} \text{a) } \lim_{n \rightarrow \infty} [n] &= (1 - q)^{-1} & \text{b) } \sum_{j=0}^{\infty} \frac{1}{[m]^{sj}} &= \frac{[m]^s}{[m]^s - 1} \\ \text{c) } \sum_{j=1}^{\infty} \frac{1}{[m]^{sj}} &= \frac{1}{[m]^s - 1}. \end{aligned}$$

All three parts readily follow from the standard formula for summing convergent geometric series.

Since each term in the sum $\zeta_q(s)$ is positive, it suffices to show that the sum of any rearrangement of these terms converges. Towards this end, suppose that p_k denotes the k th smallest prime. Let

$$\mathfrak{p}(k) = \{n \in \mathbb{Z}^+ : p_k \text{ is the largest prime divisor of } n\}.$$

Since the union $\bigcup_{k=1}^{\infty} \mathfrak{p}(k)$ consists of pair-wise disjoint sets and is equal to $\mathbb{Z}^+ \setminus \{1\}$, the sum

$$S = 1 + \sum_{k=1}^{\infty} \sum_{n \in \mathfrak{p}(k)} \frac{q^{s v(n)}}{\{n\}^s}$$

is a rearrangement of $\zeta_q(s)$.

Clearly, $v(n) \geq k - 1$ for all $n \in \mathfrak{p}(k)$. Furthermore, from the definition of $\mathfrak{p}(k)$ and from parts (b) and (c) of Lemma 2, it follows that

$$\begin{aligned} \sum_{n \in \mathfrak{p}(k)} \frac{1}{\{n\}^s} &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \cdots \sum_{m=0}^{\infty} \sum_{r=1}^{\infty} \frac{1}{[2]^{is} [3]^{js} \cdots [p_{k-1}]^{ms} [p_k]^{rs}} \\ &= \frac{[2]^s}{[2]^s - 1} \frac{[3]^s}{[3]^s - 1} \cdots \frac{[p_{k-1}]^s}{[p_{k-1}]^s - 1} \frac{1}{[p_k]^s - 1}. \end{aligned}$$

For convenience, let c_k denote this sum. Together, the comments of this paragraph imply that

$$S \leq 1 + \sum_{k=1}^{\infty} q^{s(k-1)} \sum_{n \in \mathfrak{p}(k)} \frac{1}{\{n\}^s} = 1 + \sum_{k=0}^{\infty} c_{k+1} q^{sk}.$$

The proof will be complete if the last series on the above right can be shown to converge. This is easy. Since $0 < q < 1$ and $s > 1$, it follows from the definition of c_k and from Lemma 2(a) that

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{c_{k+2} q^{(k+1)s}}{c_{k+1} q^{ks}} &= \lim_{k \rightarrow \infty} q^s \frac{[p_{k+1}]^s}{[p_{k+2}]^s - 1} = q^s \frac{(1 - q)^{-s}}{(1 - q)^{-s} - 1} \\ &= \frac{q^s}{1 - (1 - q)^s} < \frac{q^s}{1 - (1 - q)} = q^{s-1} < 1. \end{aligned}$$

The ratio test for convergence now implies the desired conclusion.

7. A PROOF OF THEOREM 2. As Theorem 2 may be established in a manner completely analogous to the way in which Euler's product formula for $\zeta(s)$ is proven, only a brief outline of its proof will be given.

Assume that $0 \leq q \leq 1$ and that $s > 1$. For $m \in \mathbb{Z}^+$, let

$$Q_m = \sum_{p \leq m} \left(1 - \frac{q^{s(p-1)}}{[p]^s} \right)^{-1}.$$

Because $0 < q^{s(p-1)}/[p]^s < 1$ for all primes p , each factor of Q_m may be expanded into an absolutely convergent series:

$$\left(1 - \frac{q^{s(p-1)}}{[p]^s} \right)^{-1} = \sum_{k=0}^{\infty} \frac{q^{sk(p-1)}}{[p]^{sk}}.$$

Moreover, viewed as a finite product of absolutely convergent series, Q_m may also be expressed as a convergent series. In fact,

$$Q_m = \prod_{p \leq m} \left(\sum_{k=0}^{\infty} \frac{q^{sk(p-1)}}{[p]^{sk}} \right) = \sum_1 \frac{q^{s\nu(n)}}{\{n\}^s}$$

where \sum_1 is summed over the set of natural numbers n having no prime factors greater than m . Comparing Q_m with $\zeta_q(s)$ leads to the inequality

$$|\zeta_q(s) - Q_m| \leq \sum_2 \frac{q^{s\nu(n)}}{\{n\}^s}$$

where \sum_2 is summed over all n having a prime factor greater than m . By Theorem 1, \sum_2 goes to 0 as $m \rightarrow \infty$. Thus, Q_m converges to $\zeta_q(s)$.

8. CONCLUDING REMARKS. Many special functions (including the exponential, gamma, beta, and the classical orthogonal polynomials of Jacobi, Legendre, Laguerre, and Hermite) have well-known q -analogs. The book by George Gasper and Mizan Rahman [1] contains an excellent exposition on special q -functions from the perspective of basic hypergeometric series. Of interest from the standpoint of probability is a recent article by Gilbert Labelle [2]: He considers a q -analog of Euler's gamma function within the context of a geometrical distribution. The extent to which classical functional equations involving both the zeta and gamma functions have well-behaved q -analogs appears to be an open question.

There are also several Bernoulli schemes in which the probability of tails occurring may be directly identified with well-known q -analogs in combinatorics. A survey of such schemes may be found in [3].

REFERENCES

1. G. Gasper and M. Rahman, *Basic Hypergeometric Series*, Cambridge Univ. Press, 1990.
2. G. Labelle, A propos d'un q -analogue pour la fonction gamma d'Euler, *Ann. Sc. Math. Québec*, 1982, vol. 6, no. 2, pp. 163–196.
3. D. P. Rawlings, Bernoulli trials and permutation statistics, *Internat. J. Math. & Math. Sc.*, 1992, vol. 15, no. 2, pp. 291–312.
4. D. P. Rawlings, Limit formulas for q -exponential functions, *Discrete Math.* 126 (1994) 379–383.

Mathematics Department
California Polytechnic State University
San Luis Obispo, CA 93407
drawling@math.calpoly.edu

What Is the Shape of a Mylar Balloon?

William H. Paulsen

Mylar balloons of different shapes and sizes have become popular as gifts or in bouquets. The most common balloons are comprised of two circular sheets of mylar, fused together at the circumference. A small opening on the circumference allows the balloon to be inflated with either air or helium. These balloons are not spherical, which is at first surprising, for it is well known that the sphere gives the maximal volume for a given surface area. Thus, these balloons suggest the following mathematical problem: Given a circular mylar balloon of deflated radius r , what will be the shape of the balloon when it is fully inflated? In particular, we could ask

- 1) What is the radius of the inflated balloon?
- 2) What is the thickness of the inflated balloon?
- 3) What is its volume?

In this paper, we will answer all three of these questions.

We begin by forming a mathematical model of the balloon. We will ignore the small opening in the balloon which is used to inflate the balloon, so the deflated balloon will have circular symmetry. However, as we inflate the balloon, crimping occurs near the “equator” of the balloon. From this, it is apparent that the surface area is not the constraining factor for the volume. To understand what is constraining the volume, let us look at a cross sectional view through the center of the balloon.

Let a be the radius of the inflated balloon. We let $y = f(x)$ describe the curve of the cross section in the first octant, as in Figure 1. Because the mylar does not stretch by a significant amount, we have that the length of the curve $f(x)$ from 0 to a is constrained by the original radius r . That is,

$$\int_0^a \sqrt{1 + f'(x)^2} dx = r. \quad (*)$$

Thus, even though the actual mylar balloon has wrinkles, the basic shape, or convex hull, of the balloon will be determined by the constraint (*).

We will assume that these wrinkles do not affect the volume of the balloon significantly, so the total volume, by the shell method, is given by

$$V = 4\pi \int_0^a xf(x) dx.$$

We want to find $f(x)$ and a such that V is maximized subject to the subsidiary condition (*) and the end point condition $f(a) = 0$.

Of course, the crimping will greatly affect the surface area of the balloon, which in practice should remain constant as the balloon is inflated. Once the shape of the

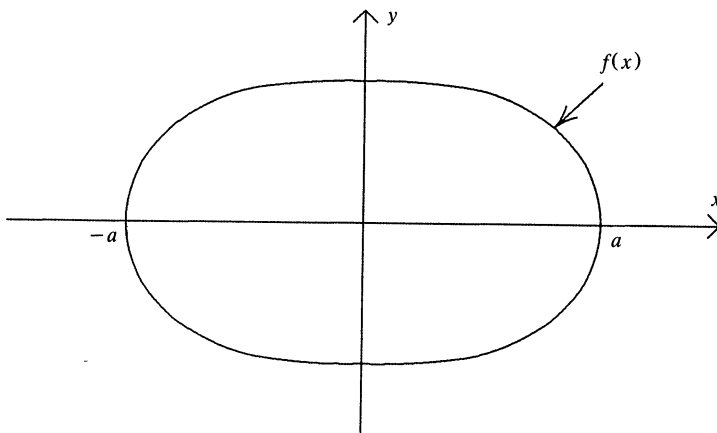


Figure 1

balloon is determined, we will be able to determine just how much crimping is taking place at different points of the balloon.

We proceed using the technique of the calculus of variations. In general, the extremum function $y(x)$ to the functional

$$J[y] = \int_a^b F(x, y, y') dx,$$

with fixed boundary conditions $y(a) = A$ and $y(b) = B$ satisfies Euler's equation [3, p. 184]

$$F_y - \frac{d}{dx} F_{y'} = 0.$$

In the current setting, we also have a subsidiary condition (*) and neither of the two end points are fixed. We can work the problem by first considering the endpoints as fixed, and then use the transversality conditions at the endpoints to determine a and $f(0)$.

To incorporate the subsidiary condition, we will use a technique not unlike the method of Lagrange multipliers. That is, the extremum function of $\int_a^b F(x, y, y') dx$ subject to the constraint $\int_a^b G(x, y, y') dx = l$ will also be an extremum of the functional

$$\int_a^b (F + \lambda G) dx$$

for some constant λ .

In the current setting, $F(x, y, y') = xy$, and $G(x, y, y') = \sqrt{1 + (y')^2}$. Thus, we seek a maximal function $y = f(x)$ of the functional

$$\int_0^a (xy + \lambda \sqrt{1 + (y')^2}) dx.$$

By Euler's equation, $y = f(x)$ satisfies

$$x - \lambda \frac{d}{dx} y' (1 + (y')^2)^{-1/2} = 0, \quad \text{or}$$

$$\lambda \frac{y'}{\sqrt{1 + (y')^2}} = \frac{x^2}{2} + C$$

for some constants λ and C .

To find the value of these two constants, we note that the volume must be maximized over all choices of a and $f(0)$. Thus, we must come up with the transversality conditions for the endpoints. In this case, it is easier to argue the transversality conditions geometrically. Because of the nature of the problem, the largest volume will be obtained when $f(x)$ is orthogonal to both axes [2, p. 61]. That is,

$$f'(0) = 0 \quad \text{and} \quad \lim_{x \rightarrow a^-} f'(x) = -\infty.$$

By plugging in $x = 0$ into the above equation, and using the condition $f'(0) = 0$, gives us $C = 0$. Thus,

$$\frac{f'(x)}{\sqrt{1 + f'(x)^2}} = \frac{x^2}{2\lambda}.$$

Since we want $f'(x) < 0$ for $0 < x < a$, let us write $2\lambda = -k^2$. Then, solving for $f'(x)$, we get

$$f'(x) = \frac{-x^2}{\sqrt{k^4 - x^4}}.$$

Since we want $f(a) = 0$, we can write the solution as

$$f(x) = \int_x^a \frac{t^2}{\sqrt{k^4 - t^4}} dt.$$

This is an elliptic integral, which has no closed form solution. But in spite of its complexity, we will still be able to answer some questions.

Let us begin by solving for k and a . The volume will be the largest when a is chosen such that the tangent at $(0, a)$ is vertical. That is, $\lim_{x \rightarrow a^-} f'(x) = -\infty$. This will happen if $a = k$. Thus,

$$f(x) = \int_x^a \frac{t^2}{\sqrt{a^4 - t^4}} dt, \quad (0 \leq x \leq a).$$

The graph of $f(x)$ is given in Figure 1. To find a , we will use the constraint (*). Since $f'(x) = -x^2 / \sqrt{a^4 - x^4}$, this becomes

$$\int_0^a \frac{a^2 dx}{\sqrt{a^4 - x^4}} = r.$$

This is another elliptic integral, but it is a definite elliptic integral, which we can put in closed form using a substitution. Let $x = at^{1/4}$, so $dx = at^{-3/4}/4 dt$. The integral becomes

$$\frac{a}{4} \int_0^1 t^{-3/4} (1-t)^{-1/2} dt = r.$$

This is a *Beta integral* [1, p. 288]. The general form of the Beta integral is

$$\int_0^1 x^m (1-x)^n dx = \frac{\Gamma(m+1)\Gamma(n+1)}{\Gamma(m+n+2)}, \quad (m > -1, n > -1).$$

Here, $m = -3/4$, and $n = -1/2$. Thus,

$$r = \frac{\Gamma(1/4)\Gamma(1/2)a}{4\Gamma(3/4)}.$$

Since $\Gamma(p)\Gamma(1-p) = \pi/\sin(p\pi)$ for $0 < p < 1$, we have that $\Gamma(1/2) = \sqrt{\pi}$ and $\Gamma(3/4) = \pi\sqrt{2}/\Gamma(1/4)$. Thus,

$$a = \frac{4r\sqrt{2\pi}}{\Gamma(1/4)^2}.$$

Hence, the radius of the inflated balloon is about 0.7627 times the radius of the deflated balloon.

The thickness of the balloon can be determined as $2f(0)$, or

$$2\int_0^a \frac{x^2 dx}{\sqrt{a^4 - x^4}}.$$

Making the same substitution as before, we get

$$\frac{a}{2} \int_0^1 t^{-1/4}(1-t)^{-1/2} dt = \frac{\Gamma(3/4)\Gamma(1/2)a}{2\Gamma(5/4)}.$$

We can use the fact that $\Gamma(p+1) = p\Gamma(p)$ for all p to simplify $\Gamma(5/4) = \Gamma(1/4)/4$. Thus, the thickness of the balloon is given by

$$\frac{16\pi^2 r}{\Gamma(1/4)^4},$$

or about 0.9139 times the radius of the deflated balloon.

Finally, the volume of the balloon is given by

$$V = 4\pi \int_0^a x f(x) dx = 4\pi \int_0^a \int_x^a \frac{xt^2}{\sqrt{a^4 - t^4}} dt dx.$$

If we treat this as a double integral over the region $0 < x < a$ and $x < t < a$, we can change the order of integration to get

$$V = 4\pi \int_0^a \int_0^t \frac{xt^2}{\sqrt{a^4 - t^4}} dx dt = \int_0^a \frac{2\pi t^4}{\sqrt{a^4 - t^4}} dt.$$

Using $t = au^{1/4}$, we get

$$V = \int_0^1 a^3 \frac{\pi}{2} u^{1/4} (1-u)^{-1/2} du = \frac{a^3 \pi \Gamma(5/4) \Gamma(1/2)}{\Gamma(7/4)} = \frac{64\pi^2 r^3}{3\Gamma(1/4)^4}.$$

Thus, the volume of the inflated balloon is about $1.2185r^3$, where r is the radius of the deflated balloon.

At this point we can estimate how much crimping will take place on different parts of the balloon. To do this, consider the effect on the balloon if the mylar was able to shrink, but not stretch. The overall shape of the balloon would remain unchanged, but the crinkles would be able to be smoothed out, forming the convex shape given by $f(x)$, which I will call the *rectified* shape of the balloon. Of course, the true shape of the balloon is different, since the mylar does not shrink.

Let us define the *crimping factor* at a point on the balloon to be the ratio between the surface area of a small patch of the balloon to the corresponding

patch on the rectified balloon, taken in the limit as the area of the patch approaches 0. By this definition, the crimping factor must be greater or equal to 1. Since we have a model of the rectified balloon, given by $f(x)$, we can use this to compute the surface area of a small patch, and compare this to the same patch of the deflated balloon.

Consider the patch of the inflated balloon which lies over the polar interval $x < r < x + \Delta x$, $\alpha < \theta < \alpha + \Delta\alpha$. The surface area of this patch of the rectified balloon is given by

$$\int_x^{x+\Delta x} \Delta\alpha s \sqrt{1 + f'(s)^2} ds \approx \Delta x \Delta\alpha x \sqrt{1 + f'(x)^2} = \frac{xa^2 \Delta x \Delta\alpha}{\sqrt{a^4 - x^4}}.$$

Note that a point x units from the center of the inflated balloon came from a point

$$\int_0^x \sqrt{1 + f'(s)^2} ds = \int_0^x \sqrt{1 + \frac{s^4}{a^4 - s^4}} ds = \int_0^x \frac{a^2 ds}{\sqrt{a^4 - s^4}}$$

units away from the center of the deflated balloon. Thus, the same patch on the deflated balloon would be the polar interval

$$\int_0^x \frac{a^2 ds}{\sqrt{a^4 - s^4}} < r < \int_0^{x+\Delta x} \frac{a^2 ds}{\sqrt{a^4 - s^4}}, \quad \alpha < \theta < \alpha + \Delta\alpha.$$

The area is given approximately as

$$\frac{a^2 \Delta x \Delta\alpha}{\sqrt{a^4 - x^4}} \int_0^x \frac{a^2 ds}{\sqrt{a^4 - s^4}}.$$

Thus, the ratio between these two areas is

$$C(x) = \frac{1}{x} \int_0^x \frac{a^2 ds}{\sqrt{a^4 - s^4}}.$$

Note that $\lim_{x \rightarrow 0} C(x) = 1$, so there is no crimping at the “poles” of the balloon. The maximum crimping occurs at the equator, $x = a$, for which $C(x) = r/a = \Gamma(1/4)^2 / 4\sqrt{2}\pi \approx 1.32$. For comparison, this is about the same amount of crimping as the function $1.23 \sin(x)$.

It is interesting to compare the mathematical results to some experimental measurements. Although the volume of a mylar balloon is difficult to measure, the radii for the inflated and deflated balloons, as well as the thickness, are relatively easy. Using the standard technique of measuring five times, and averaging the middle three measurements, I found that the deflated diameter of a given mylar balloon was 43.85 cm, while the diameter of the inflated balloon was 33.35 cm. The thickness was measured to be 20.166 cm. Thus, $r \approx 21.925$ cm, $a \approx 16.675$ cm, and $f(0) \approx 10.083$ cm. The ratio of the radii is experimentally 0.7605, which is surprisingly close to the theoretical result of 0.7627. The ratio of the thickness of the balloon to the original radius is experimentally 0.9198, compared to the theoretical result of 0.9139.

This problem could be generalized by considering different shapes of balloons. That is, instead of starting with two circular disks, we fuse together two pieces of mylar of some other shape. The most common variant would be the heart-shaped balloons, seen mostly around Valentine’s Day. However, this problem would be very formidable analytically. A more reasonable problem would be to find the

shape of a toroid balloon, formed by two annular rings. This shape is suggested by the inflatable swimming rings. This problem could be solved using similar methods as above, except the boundary conditions would cause the problem to be much more complex.

REFERENCES

1. William H. Beyer, Editor, *CRC Standard Mathematical Tables*, 28th edition, CRC Press (Boca Raton, Florida) 1987.
2. I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, Prentice-Hall, Inc. (Englewood Cliffs, N. J.) 1963.
3. Jerry B. Marion, *Classical Dynamics of Particals and Systems*, second edition, Academic Press (New York) 1970.

Department of Computer Science, Mathematics, & Physics
Arkansas State University
P. O. Box 70
State University, AR 72467-0070
wpaulsen@quapaw.astate.edu

The Law and Mathematicians

From an article by Henry Louis Gates, Jr., in *ACADEME* (January-February, 1994, p. 17):

One recalls Justice William O. Douglas's 1973 remarks, "One of the most offensive experiences in my life was a visit to anation where bookstalls were filled only with books on mathematics and books on religion."

Submitted by Raymod Greenwell
Hofstra University

From a letter by Doug Jungreis (UCLA):

I was on jury duty for 3 weeks... I was assigned to a few jury selections but never put on the jury. In one jury selection, the third question that the judge asked all the potential jurors was, "Are you or a family member a mathematican or engineer?" Apparently they have trouble with us math people because we don't understand the idea of "proof beyond a reasonable doubt."

Submitted by Joe Gallian

Euler's Theorem for Polyhedra: A Topologist and Geometer Respond

Peter Hilton and Jean Pedersen

In their stimulating paper [1], to which we here make a friendly and constructive response, the authors, Branko Grünbaum and Geoffrey Shephard, introduce the interesting geometric concept of *polyhedral set*, generalizing the familiar notion of polyhedron, but confining themselves, for their present purposes, to subsets of \mathbb{R}^3 . They discuss dissections of such sets, especially relatively open convex dissections of bounded polyhedral sets and show, by easily accessible arguments, the nice properties of the Euler characteristic χ relative to such dissections. They are thereby led to a formula for $\chi(P)$, namely, Theorem 4 of [1],

'If P is a polyhedral set, then $V - E + F - C = \chi(P)$.'

Here $V, -E, F, -C$ are defined as χ (k -scaffold of P), $k = 0, 1, 2, 3$, so the resemblance of their Theorem 4 to the classical theorem of Euler might seem at first to owe more to a judicious choice of notation than to a conceptual closeness; however, careful inspection of their definition of a k -scaffold and the contemplation of examples show that their notation may, indeed, be justified.

The paper is, in our judgment, a very real contribution to polyhedral geometry, clearly written, and enriched, as one would expect of these authors, by clear illustrative diagrams. Why then did we seek the approval of the Editor of the *Monthly* to publish a response? It is because the paper appears, in contrast to its obvious positive contribution, to both neglect and distort the contribution which algebraic and combinatorial topologists have made to the development of the Euler characteristic. We claim that the 'problem', referred to on page 110 of [1], that 'a simple numerical identity¹ is now seen to be hedged with additional conditions or exceptional cases' is no real problem at all, and that topologists have completely and satisfactorily formalized the Euler characteristic (now generalized to the Euler-Poincaré characteristic and applied to compact polyhedra of any dimension) and in the process fully understood its topological—and hence combinatorial—invariance. Moreover, we deny categorically that 'this development led to a loss of the connection to the origins of Euler's theorem as a relation involving the vertices, edges and faces of a polyhedral object.' It is an irony that this extraordinary allegation should be almost the only reference to topological developments in 5 pages devoted to 'Historical Remarks and Comments.'

Specifically, we reject the calculation of $v - e + f$ for Figure 1(b) which we here reproduce.² We claim that a polyhedron is made of cells and a closed n -cell is the homeomorph of the n -dimensional ball $x_1^2 + x_2^2 + \cdots + x_n^2 \leq 1$ in \mathbb{R}^n . In

¹The authors of [1] are, of course, referring to $v - e + f = 2$.

²Reproduced figures are accompanied by their original captions.

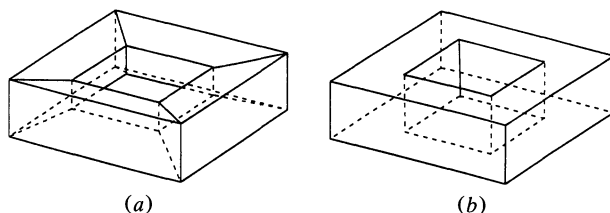


Figure 1. (a) A polyhedron to which Euler's theorem in its elementary form $v - e + f = 2$ does not apply. (b) A polyhedron of genus $g = 1$ to which Euler's theorem in the form $v - e + f = 2 - 2g$ does not apply.

particular, a 2-dimensional face must be a polygonal region. Indeed, one might legitimately comment that it is unreasonable to expect that one can count the number of holes by counting subsets which are themselves allowed to have holes! Moreover, if one allows strange (2-dimensional) faces as in Figure 1(b), then why not regard the entire surface as a face and thus obtain $v - e + f = 1$ for *any* surface! Finally, we remark that it should never happen that scratching a few extra line segments on a surface—the only difference between Figures 1(a) and 1(b) of [1]—should alter a really significant invariant quantity.

The error in permitting Figure 1(b) as a legitimate dissection of the torus is to be found in many parts of the world (including, for example, the UK and New Zealand); so much so that we have ourselves written an article [2] attempting to stem the tide of error.

We regard the theorem (referred to somewhat disparagingly in [1]) that the Euler-Poincaré characteristic equals the alternating sum of the Betti numbers as fundamental. It not only explains Euler's original result (since orientable closed surfaces bounding convex regions are homeomorphic to the sphere), but establishes the *homotopy* invariance (not merely the topological invariance) of χ . Of course it also explains formula (2) of [1], that is

$$v - e + f = 2 - 2g, \quad (2)$$

for an orientable closed surface of genus g , since the Betti numbers of such a surface are given by $p_0 = 1$, $p_1 = 2g$, $p_2 = 1$. Thus formula (2) is scarcely a great stride in extending Euler's Theorem and so should not be seen as 'dealing with a discrepancy'. Nor should Figure 2 of [1] create a problem (once one has understood the instructions for the construction of the polyhedron P in question³; apparently the top and bottom faces of the cube are missing and the 'drilling' takes place through empty space—are we right?). Certainly $\chi(P) = 1$ for this polyhedron, so formula (2) does not apply. But P is not a closed orientable surface, so how could formula (2) apply?

In this century topologists have made great strides generalizing the properties of the Euler-Poincaré characteristic (the Hopf Trace Formula, the Lefschetz Fixpoint Theorem, ...), and applying it to many problems like the vector field problem and studies of multiplicative structures on manifolds. We may also cite Gottlieb's Theorem on aspherical finite complexes X : If $\chi(X) \neq 0$, then the fundamental group of X has trivial center. Very recently, Ross Geoghegan and Andrew Nicas

³We interpret this figure as obtained from a sphere by identifying two of its points; of course, the resulting space is not a manifold at this point.

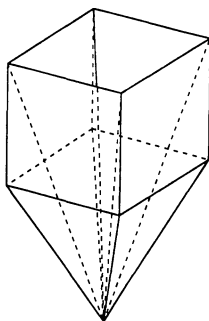


Figure 2. A funnel-shaped polyhedron for which the elementary forms of Euler's theorem are not valid.

have been engaged in developing important higher-order analogs χ_n of χ which are relevant, for example, to the study of homotopies of self-maps of manifolds. They further show essentially that, if X is aspherical, then $\chi_1(X) \neq 0$ implies that the fundamental group of X is infinite cyclic.

It is splendid that geometers as well as topologists are making progress in this field; let neither disparage the efforts and successes of the other.

REFERENCES

1. Branko Grünbaum and G. C. Shephard, A new look at Euler's theorem for polyhedra, *Amer. Math. Monthly*, 101 (1994), 109–128.
2. Peter Hilton and Jean Pedersen, A deplorable fallacy—and a minor fault, *Math. Gazette* (1994), 286–289.

Department of Mathematical Sciences
State University of New York
Binghamton, NY 13902-6000

Department of Mathematics
Santa Clara University
Santa Clara, CA 95053

RESPONSE FROM GRÜNBAUM AND SHEPHARD

Our paper *A New Look at Euler's Theorem for Polyhedra* seems to have caused some lively discussion. Naturally we are very pleased; it is far better for a paper to attract comment (even if critical) than to gather dust, unread, on the shelves of libraries.

However, the “Response” by Peter Hilton and Jean Pedersen seems to us to be misleading and inappropriate. On a matter of detail, we never claimed that the “calculation” of the Euler characteristic of the polyhedron in Figure 1(b) was correct. It was presented as an example of how earlier authors, hampered by inadequate definitions, were misled in their treatment of the subject. Our intention is clear from the “Introduction”, as well as from the more detailed discussion in the “Historical Remarks and Comments.” The rejection of that “calculation”, and of the similar one for the polyhedron in Figure 2, was our explicit aim, obviously missed by Hilton and Pedersen.

However, our objection to the “Response” is much deeper. Their comment that “topologists have completely and satisfactorily formalized the Euler characteristic (now generalized to the Euler-Poincaré characteristic as applied to compact

polyhedra of any dimension)" is true, but irrelevant. In particular, we reject the implication that the geometrical approach we propose is redundant. The purpose of our paper, as stated in the Introduction, is to show how the "traditional" approach of counting vertices, edges and faces, can, with care and appropriate modifications, lead to results which are valid and far-reaching. In particular, there is no need, as topologists generally do, to restrict attention to bounded polyhedra which can be expressed as finite unions of topological balls all of which are open, or all of which are closed. This restriction eliminates many of our examples in the paper, to which our methods apply. In particular, it seems unreasonable to us to reject a polyhedron such as that shown in Figure 9(a) by stating dogmatically that every "2-dimensional face must be a polygonal region" and then to define a polygonal region in such a way as to eliminate multiply-connected regions. Euler himself, if he were alive today, would, we believe, feel far more comfortable with our approach to his theorem than that advocated in the "Response".

Our purpose in writing the paper was to show that a simple geometrical treatment of a classical topic can lead to results that are more general, and more intelligible to the majority of mathematicians, than the recondite and much more abstract topological approach. Hence our "neglect" of topology. On the other hand, we would be very pleased if topological generalizations of our approach were to be developed.

We would like to take this opportunity to point out some errors that crept into the captions of two of the figures in our paper; we are grateful to many colleagues for drawing our attention to these. In Figure 4, we should have: (c) 0 (8, 8, 0); (d) 6 (8, 0, -2); (f) 4 (6, 2, 0). In Figure 5, the correct values are: (c). -4 (3, 9, 2); (d) 2 (2, 5, 5).

(Added August 19, 1994) In "Die Euler-Poincaré- Charakteristik und ihre topologische Weiterentwicklung" (Elemente der Mathematik, 49(1994), pp. 66–76) P. J. Hilton refers to an unnamed paper, which appears to be our paper in the "Monthly". He presents there objections very similar to those in the above Response—in fact he goes further and describes parts of the paper as "complete nonsense!" ("vollkommener Unsinn!").

*Department of Mathematics, GN-50
University of Washington
Seattle, WA 98195
grunbaum@math.washington.edu*

*School of Mathematics
University of East Anglia
Norwich NR4 7TJ, England
G.Shephard@uea.ac.uk*

Common sense is the collection of prejudices acquired by age eighteen.

—*Albert Einstein* (1879–1955)

E T Bell, *Mathematics, Queen and Servant of the Sciences*,
New York: McGraw Hill, 1951.

The Role of Paradoxes in the Evolution of Mathematics

I. Kleiner and N. Movshovitz-Hadar

A paradox has been described as a truth standing on its head to attract attention. Undoubtedly, paradoxes captivate. They also cajole, provoke, amuse, exasperate, and seduce. More importantly, they arouse curiosity, stimulate, and motivate.

In this paper we present examples of paradoxes from the history of mathematics which have inspired the clarification of basic concepts and the introduction of major results. Our examples will deal with numbers, logarithms, functions, continuity, tangents, infinite series, sets, curves, and decomposition of geometric objects.

We will use the term “paradox” in a broad sense to mean an inconsistency, a counterexample to widely held notions, a misconception, a true statement that seems to be false, or a false statement that seems to be true. It is in these various senses that paradoxes have played an important role in the evolution of mathematics. Indeed, as Bell and Davis, respectively, put it:

The mistakes and unresolved difficulties of the past in mathematics have always been the opportunities of its future ([1], p. 283).

One of the endlessly alluring aspects of mathematics is that its thorniest paradoxes have a way of blooming into beautiful theories ([6], p. 55).

Paradoxes can also serve a useful role in the classroom. The temporary confusion and insecurity which they may engender in students can be put to good use. Conflict and predicament are useful pedagogical devices (provided, of course, that they are dealt with). They may foster a positive attitude to “getting stuck,” provide the opportunity to participate in debate and controversy over mathematical issues, and promote the realization that mathematics often develops in this very way. Teachers may gain a better appreciation of students’ difficulties in coming to grips with concepts and results with which some of the greatest mathematicians of all time struggled. Such concepts and results, while paradoxical and challenging at the time, became commonplaces in subsequent generations. In the words of Kasner and Newman ([12], p. 193):

The testament of science is so continually in a flux that the heresy of yesterday is the gospel of today and the fundamentalism of tomorrow.

PARADOXES INVOLVING NUMBERS. The evolution of the concept of number has been beset by paradoxes almost every step of the way. As P. J. Davis put it ([7], p. 305):

It is paradoxical that while mathematics has the reputation of being the one subject that brooks no contradictions, in reality it has a long history of successful living with contradictions. This

is best seen in the extensions of the notion of number that have been made over a period of 2500 years. From limited sets of integers, to fractions, negative numbers, irrational numbers, complex numbers, transfinite numbers, each extension, in its way, overcame a contradictory set of demands.

The first sentence in the above quotation may be thought of as a “metaparadox”—a nontechnical, paradoxical statement about technical, paradoxical phenomena. We will point out a variety of such metaparadoxes; they are interesting in their own right as issues for philosophical discussion or contemplation. But now to some paradoxes dealing with the evolution of various number systems.

(a) The Pythagoreans of the 6th century B.C. believed that every line segment can be measured by a positive integer or the ratio of two such integers. This was to them not merely a very plausible fact, but an article of faith, an aspect of their philosophy. Moreover, the idea formed the basis of the pythagorean theory of proportion (see [23]). It was thus a great shock (paradox) to them when they discovered that the diagonal of a unit square cannot be measured by a whole number or by a ratio of whole numbers; or, as the Greeks put it, that the diagonal and side of a square are *incommensurable*. Their proof of this result is essentially the one we use today to show that $\sqrt{2}$ is irrational. The paradox was arrived at by using the Pythagorean Theorem. Thus the

Metaparadox: The Pythagorean Theorem was the undoing of the pythagorean philosophy and the pythagorean theory of proportion.

The discovery of the incommensurability of the diagonal and side of a square had far-reaching consequences for Greek mathematics. On the positive side, it inspired Eudoxus to found a sophisticated theory of proportion which applied to both commensurable and incommensurable magnitudes. This, in turn, motivated Dedekind more than two millennia later to define the real numbers via Dedekind cuts. On the debit side, it turned the direction of Greek mathematics (at least in its very productive, classical period) from a harmonious collaboration of number and geometry to an almost exclusive concern with geometry.

(b) The introduction of negative numbers into mathematics and their subsequent use occasioned considerable consternation and difficulties. A major conceptual framework that had to be abandoned was the prohibition of subtracting a greater from a smaller number. As Wallis in the 17th century put it ([20], p. 438): “[How can] any magnitude... be less than nothing, or any number fewer than none?”

Among other paradoxes having to do with negative numbers are the following two:

(i) Wallis “proved” that negative numbers are greater than infinity. His argument was that since (for positive a) $\frac{a}{0} = \infty$, $a/\text{a neg. no.} > \infty$; this is so because decreasing the denominator increases the fraction.

(ii) In a letter to Leibniz, Arnauld (a 17th-century mathematician and philosopher) objected to the equality $\frac{1}{-1} = \frac{-1}{1}$ on the grounds that the ratio of a greater to a smaller quantity cannot equal the ratio of a smaller to a greater. Leibniz agreed this was a difficulty, but argued for the tolerance of negative numbers because they are useful and, in general, lead to consistent results. See [5], pp. 39–40.

Justification of otherwise inexplicable notions on the grounds that they yield useful results has occurred frequently in the evolution of mathematics. This brings up the following

Metaparadox: How can meaningless (or at best inexplicable) things be so useful?

Of course, out of meaninglessness (or confusion) emerged, in time, clarity and understanding.

(c) The solution by radicals of cubic equations was one of the great achievements of 16th-century mathematics. Cardan's solution of the cubic $x^3 = ax + b$ was given by the formula

$$x = \sqrt[3]{\frac{b}{2} + \sqrt{\left(\frac{b}{2}\right)^2 - \left(\frac{a}{3}\right)^3}} + \sqrt[3]{\frac{b}{2} - \sqrt{\left(\frac{b}{2}\right)^2 - \left(\frac{a}{3}\right)^3}}.$$

Bombelli applied it to the equation $x^3 = 15x + 4$ to obtain $x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}$. Cardan had earlier denied the applicability of his formula to such equations since it introduced square roots of negative numbers, which he rejected. But Bombelli noted (by inspection) that $x = 4$ is a solution of $x^3 = 15x + 4$. (The other two roots, $-2 \pm \sqrt{3}$, are also real.) Here was a paradox: The roots of $x^3 = 15x + 4$ are real, yet the formula yielding the roots involved complex, and at the time meaningless, numbers. "The whole matter seemed to rest on sophistry rather than on truth," noted Bombelli ([15], p. 2). And he set himself the task of resolving that sophistry, which resulted in the birth of complex numbers.¹ Birth, however, did not entail legitimacy. It took another two and a half centuries before complex numbers were accepted as bona fide mathematical entities.

PARADOXES INVOLVING LOGARITHMS. The issue of the meaning of logarithms of negative and complex numbers arose in the early 18th century in connection with integration. In analogy with the real case, Johann Bernoulli integrated $1/(x^2 + a^2)$ as follows:

$$\begin{aligned} \int \frac{dx}{x^2 + a^2} &= \int \frac{dx}{(x + ai)(x - ai)} = -\frac{1}{2ai} \int \left(\frac{1}{x + ai} - \frac{1}{x - ai} \right) dx \\ &= -\frac{1}{2ai} [\log(x + ai) - \log(x - ai)] = -\frac{1}{2ai} \log \frac{x + ai}{x - ai}. \end{aligned}$$

In an exchange of letters (begun in 1702 and lasting sixteen months) Bernoulli and Leibniz argued about the meaning of $\log(x + ai/x - ai)$, and, in particular, about the meaning of $\log(-1)$. Bernoulli asserted that $\log(-1)$ is real while Leibniz claimed it is imaginary, each advancing various arguments to support his view. For

¹Bombelli developed rules for manipulating expressions of the form $a + b\sqrt{-1}$ and was thereby able to show that (one of the values of) $\sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}$ is indeed 4.

example, Bernoulli argued that since

$$\frac{dx}{x} = \frac{d(-x)}{-x}, \quad \int \frac{dx}{x} = \int \frac{d(-x)}{-x},$$

hence $\log x = \log(-x)$. In particular, $\log(-1) = \log 1 = 0$. Among Leibniz' arguments were the following:

(i) Since the range of $\log a$, for $a > 0$, is all real numbers, it follows that $\log a$, for $a < 0$, must be imaginary, because the real numbers have already been "spoken for".

(ii) If $\log(-1)$ were real, then $\log i$ would also be real, since $\log i = \log(-1)^{1/2} = \frac{1}{2} \log(-1)$. But this is clearly absurd, alleges Leibniz.

(iii) Putting $x = -2$ in the expansion

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots$$

yields $\log(-1) = -2 - \frac{4}{2} - \frac{8}{3} - \dots$. Since the series on the right diverges, it cannot be real, hence must be imaginary.

The above are indeed interesting examples of the art (not to say "science") of symbolic manipulation practiced by some of the greatest mathematicians of the 17th and 18th centuries. The resulting paradoxes had "for a long time... tormented me," noted Euler ([17], p. 72). He resolved them in a 1749 paper. We quote from its interesting introduction ([14], p. 4):

Since logarithms are clearly part of pure mathematics it may well be surprising to learn that they have been until now the subject of an embarrassing controversy in which whatever side is taken contradictions appear that seem completely impossible to resolve. Meanwhile if truth is to be universal there can be no doubt that these contradictions, ..., however unresolved they seem can only be apparent. ... I will bring out fully all the contradictions involved so that it may be seen how difficult it is to discover truth and to guard against inconsistency even when two great men are working on the problem.

The crux of Euler's solution was the Euler-Cotes formula $e^{i\theta} = \cos \theta + i \sin \theta$. It implies that $e^{i(\pi+2n\pi)} = \cos(\pi+2n\pi) + i \sin(\pi+2n\pi) = \cos \pi + i \sin \pi = -1$, so that $\log(-1) = i(\pi+2n\pi)$, where $n = 0, \pm 1, \pm 2, \dots$. Thus $\log(-1)$ is *multivalued* (in fact, infinite-valued) and all its values are complex. Both Bernoulli and Leibniz were wrong, the former "more so" than the latter.

PARADOXES INVOLVING FUNCTIONS. The concept of function originated in the early 18th century. Newton and Leibniz invented the calculus in the latter part of the 17th century. Here, then, is a

Metaparadox: Calculus without functions.

The calculus of Newton and Leibniz was a calculus of curves (given by equations) rather than of functions.

A function was viewed at different times as a formula, a curve, or an arbitrary correspondence. Paradoxes turned up to dethrone one or another of these views of functionality. Even the very meaning of a formula, as well as its scope (i.e., the functions that are representable by formulas), changed over time, and were often subjects of considerable controversy. For example:

(a) To Euler and his contemporaries of the mid-18th century a function meant a formula, where the latter concept, though not rigorously defined, was broadly construed to allow (among other things) infinite sums and products in its formation. There were several implicit assumptions:

(i) The function (formula) had to be given by a *single* expression. For example,

$$f(x) = \begin{cases} x, & x > 0 \\ -x, & x \leq 0 \end{cases} \text{ was not considered a function.}$$

(ii) The independent variable had to range over *all* real numbers (except possibly for isolated points, as in $f(x) = \frac{1}{x}$). For example, $f(x) = x, 0 \leq x \leq 1$, was not considered a function

(iii) Two functions which agreed on an interval were assumed to agree everywhere on the line.

The significance of these assumptions was the fact that the algorithms of the calculus applied at that time only to such functions.

Many of the 18th-century (mis)conceptions about functions were overturned by Fourier's work on heat conduction in the early decades of the 19th century. As a result of this work Fourier claimed to have shown that *any* function defined on some interval can be represented on that interval as an infinite series of sines and cosines—a *Fourier series*.² For example, if

$$f(x) = \begin{cases} -1, & -\pi < x < 0 \\ 0, & x = 0 \\ 1, & 0 < x < \pi \end{cases},$$

then

$$f(x) = \frac{4}{\pi} \left(\frac{\sin x}{1} + \frac{\sin 3x}{3} + \frac{\sin 5x}{5} + \dots \right) \text{ for all } x \in (-\pi, \pi).^3$$

Several fundamental departures concerning functions resulted from Fourier's work:

1. It became legitimate, and important, to consider functions whose domain is an *interval* rather than the entire real line.
2. Two functions could agree on an interval but differ outside the interval.
3. A function given by two or more distinct expressions could equal a function given by a single expression.

(b) In an 1829 paper on Fourier series Dirichlet introduced the so-called Dirichlet function

$$D(x) = \begin{cases} 1, & \text{if } x \text{ is rational} \\ 0, & \text{if } x \text{ is irrational.} \end{cases}$$

This function was neither a formula nor a curve. It was a new type of function, described by a correspondence. It was the first of many functions which came to be called "pathological"—but not for very long (see [25]).

²Although this result is, of course, incorrect (given *our* conception of functions) in the generality which Fourier claimed for it, a large class of functions *can* be represented by Fourier series. In fact, Fourier's contemporaries would have been hard put to find an exception.

³In the latter part of the 18th century, following debates surrounding the famous vibrating-string problem, it became legitimate (at least in some quarters) to consider functions defined by several expressions. See [16].

At the end of the 19th century Baire extended (again) the notion of formula. To him it meant an expression obtained from a variable and constants by a (possibly countable) iteration of additions, multiplications, and the taking of limits. He called such a function *analytically representable* and showed that the Dirichlet function is of this type: $D(x) = \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \cos(m! \pi x)^n$. Thus the “pathological” Dirichlet function became a “tame” analytically representable function.

Is analytic representability a universal mode of representability of functions? That is, are there functions which are not analytically representable? Yes and no. If you are a formalist, you can show by a counting argument that the set of analytically representable functions has cardinality c , while the set of all functions (clearly) has cardinality 2^c . Thus there are uncountably many functions which are not analytically representable. But no one has given a *constructive* example of even one.

PARADOXES INVOLVING CONTINUITY. Although the concept of continuity is nowadays fundamental in mathematics, its modern definition was not formulated until the 19th century, about 150 years after the invention of the calculus by Newton and Leibniz. In the 18th century, Euler did define a notion of continuity in response to the famous vibrating-string controversy ([8], p. 301). Thus a continuous function was one given by a single expression (formula), while a function given by several expressions was considered *discontinuous*. For example, to Euler the function

$$f(x) = \begin{cases} x, & x > 0 \\ -x, & x \leq 0 \end{cases}$$

was discontinuous, while the function comprising the two branches of a hyperbola was considered continuous, since it is given by the single expression $f(x) = \frac{1}{x}$ (see [16], p. 301).

The work on Fourier series showed the untenability of the 18th-century notion of continuity. For example, the function

$$g(x) = \begin{cases} -1, & -\pi < x < 0 \\ 0, & x = 0 \\ 1, & 0 < x < \pi \end{cases}$$

could (as we have seen) be represented by a single expression, namely its Fourier series, hence it was both continuous and discontinuous in the 18th-century sense of that concept.

In an 1821 work Cauchy initiated a reappraisal and reorganization of the foundations of 18th-century calculus. In this work Cauchy defined continuity essentially as we understand the concept today, although he used the then-prevailing language of infinitesimals rather than the now-accepted $\varepsilon - \delta$ formulation given by Weierstrass in the 1850s. The shift in point of view from Euler’s to Cauchy’s conceptions of continuity was fundamental. In the former case continuity was a global property while in the latter case it was a local property. But the concept of continuity proved to be very subtle, and was not completely understood even by Cauchy and his contemporaries of the early to mid-19th century. For example:

(a) Cauchy “proved” that an infinite sum (a convergent series) of continuous functions is a continuous function ([4], p. 110). This, of course, is incorrect. A

counterexample was given by Abel in the 1820s—it is essentially the series

$$\sum_{n=0}^{\infty} \frac{\sin(2n+1)x}{2n+1}$$

we encountered earlier, which is discontinuous at $x = k\pi$, $k = 0, \pm 1, \pm 2, \dots$. The error in Cauchy's proof resulted from his failure to distinguish between convergence and uniform convergence of a series of functions. In fact, "the realization of the central role of the concept of uniform convergence in analysis came about slowly in the last [19th] century" ([21], p. 97).

(b) Euler's continuous functions were, in practice, differentiable (except possibly at isolated points). So were Cauchy's—at least this is what Cauchy and his contemporaries believed, and what some of them "proved" (see [26]). It was therefore astonishing when Weierstrass in the 1860s gave an example of a continuous function which is *nowhere* differentiable, namely $f(x) = \sum_{n=1}^{\infty} b^n \cos(a^n \pi x)$, a an odd integer, b a real number in $(0, 1)$, and $ab > 1 + (3\pi/2)$. This and similar examples showed for the first time that the concept of continuity is considerably broader than that of differentiability, and thus established continuity as an important concept of investigation in its own right. The examples also showed the limitations of intuitive geometric reasoning in analysis, and thus the need for careful, analytic formulations of basic notions.

In a modern development of a different kind, Schwartz and Sobolev showed in the 1940s that every continuous function is, indeed, "differentiable". But the derivative is now a "generalized function" (a "distribution"). For example, if

$$f(x) = \begin{cases} 1, & x > 0 \\ \frac{1}{2}, & x = 0 \\ 0, & x < 0, \end{cases}$$

then

$$f'(x) = \begin{cases} 0, & x \neq 0 \\ \infty, & x = 0, \end{cases}$$

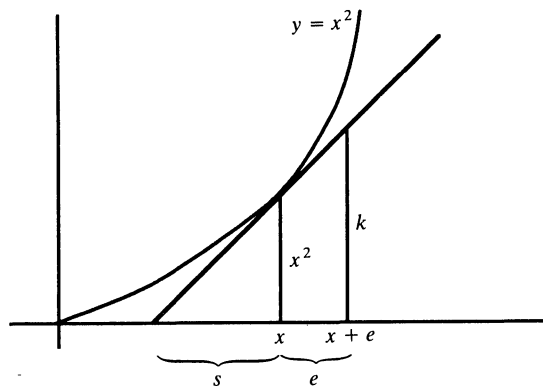
which is the Dirac delta "function" $\delta(x)$. As this example shows, there are even discontinuous functions which are differentiable (in the Schwartz/Sobolev sense)—a shocking realization (it would have been) for mathematicians of the second half of the 19th century.

PARADOXES INVOLVING ASPECTS OF THE CALCULUS (OTHER THAN CONTINUITY).

(a) The calculus was invented (independently) by Newton and Leibniz in the last third of the 17th century. But many of its important ideas were foreshadowed in early-17th century works of prominent mathematicians, notably Fermat. In the late 1630s he devised a method for dealing with problems on tangents and on maxima and minima. The following example illustrates Fermat's approach (see [8], p. 122):

Suppose we wish to find the tangent to the parabola $y = x^2$ at some point (x, x^2) . Let $x + e$ be a nearby point on the x -axis and let s denote the *subtangent* of the curve at the point (x, x^2) (see accompanying diagram). Similarity of triangles yields $x^2/s = k/(s + e)$. Fermat notes that k is "approximately equal" to $(x + e)^2$; writing this as $k \approx (x + e)^2$ we get

$$\frac{x^2}{s} \approx \frac{(x + e)^2}{s + e}.$$



Solving for s we have

$$s \approx \frac{ex^2}{(x+e)^2 - x^2} = \frac{ex^2}{x^2 + 2ex + e^2 - x^2} = \frac{ex^2}{e(2x+e)} = \frac{x^2}{2x+e},$$

hence $x^2/s \approx 2x + e$. Note that x^2/s is the slope of the tangent to the parabola at (x, x^2) . Fermat now “deletes” e and claims that the slope of the tangent is $2x$.

Fermat’s method was severely criticized by some of his contemporaries. They objected to his introduction and subsequent suppression of the mysterious e . Dividing by e meant regarding it as not zero. Discarding e implied treating it as zero. This is inadmissible, they rightly claimed. In a somewhat different context, but with equal justification, Bishop Berkeley in the 18th century would refer to such e ’s as “the ghosts of departed quantities,” arguing that “by virtue of a twofold mistake . . . [one] arrive[d], though not at a science yet at the truth” ([13], p. 428).

The justification of 17th- and 18th-century algorithms of the calculus was that they yielded correct results—another important example of the utility of “meaningless” procedures (cf. p. 965). The end seemed to have justified the means. *Rigorous* justification of the calculus—of one kind—came with the 1821 introduction of limits by Cauchy, and—of another kind—with the 1960 introduction of infinitesimals by Robinson.

Metaparadox: How can the calculus be founded on two distinct, and in some ways incompatible, theories: limits, based on the real numbers, and infinitesimals, based on the hyperreal numbers? Or, as Steen put it: “The epistemological foundation of mathematical analysis is far from settled” ([22], p. 92).

(b) Power series were a potent tool in 17th- and especially 18th-century calculus. They were manipulated as polynomials, with little if any attention paid to questions of convergence. In fact, Euler and others consciously used *divergent* series to great advantage. The results thus obtained were impressive and important, but errors and paradoxes became unavoidable. Here are two:

(i) There is undoubtedly a touch of the metaphysical in the mathematical infinite. The following example, due to Euler, confirms it ([13], p. 447): Letting $x = -1$ in $(1+x)^{-2} = 1 - 2x + 3x^2 - 4x^3 + \dots$, he gets

$$\infty = 1 + 2 + 3 + 4 + \dots (*).$$

Letting $x = 2$ in $(1 - x)^{-1} = 1 + x + x^2 + x^3 + \dots$, one has

$$-1 = 1 + 2 + 4 + 8 + \dots (**).$$

Since each term on the right side of (**) is greater than or equal to the corresponding term on the right side of (*), $-1 > \infty$. But clearly $\infty > 1$. Hence $-1 > \infty > 1$. Euler infers that ∞ must be a type of limit between the positive and negative numbers, and in this sense resembles 0.

(ii) Occasionally 17th- and 18th-century mathematicians revelled in the art of series-manipulation if for no better reason (it would seem) that to demonstrate their prowess. For example, putting $x = 1$ in $\log(1 + x) = x - x^2/2 + x^3/3 - \dots$ yields $\log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$. So far so good. But now, the argument went, the right side equals

$$\begin{aligned} & (1 + \frac{1}{3} + \frac{1}{5} + \dots) + (\frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \dots) - 2(\frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \dots) \\ &= (1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \dots) - (1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \dots) = 0, \end{aligned}$$

hence $\log 2 = 0$. It was only in the mid-19th century that Riemann resolved this paradox by proving that the sum of a conditionally convergent series can assume, upon rearrangement, *any* value. “The discovery of this apparent paradox contributed essentially to a re-examination and rigorous founding... of the theory of infinite series” ([21], p. 30).

PARADOXES INVOLVING SETS. (a) During nearly the last three decades of the 19th century Cantor developed many important set-theoretic ideas using an intuitive (“naive”) notion of set. Eventually his concept proved inadequate and led to paradoxes. Perhaps the best known is Russell’s classic paradox of 1902: Let $R = \{x: x \notin x\}$. Then $R \in R$ if and only if $R \notin R$. This paradox had a profound effect on a number of mathematicians (see [19]). It devastated the logician Frege, who had just completed a two-volume treatise on the foundations of arithmetic which relied on set-theoretic notions. Learning of Russell’s paradox, he lamented ([13], p. 1192):

A scientist can hardly meet with anything more undesirable than to have the foundation give way just as the work is finished. I was put in this position by a letter from Mr. Bertrand Russell when the work was nearly through the press.

On the other hand, the paradoxes of set theory had positive effects. In particular, they provoked mathematicians to give precise meaning to the notion of set by devising various axiomatizations of set theory (e.g., the Zermelo-Fraenkel axioms, the Russell and Whitehead theory of types, the Gödel-Bernays system). Although such axiom-systems avoided the known paradoxes, they did not guarantee that new ones would not emerge. As Poincaré put it picturesquely ([13], p. 1186):

We have put a fence around the herd to protect it from the wolves but we do not know whether some wolves were not already within the fence.

Here are two metaparadoxes resulting from Cantor’s work in set theory:

Metaparadox 1: Infinity comes in different sizes; in fact, in *infinitely many* different sizes.

The second metaparadox comes from juxtaposing the following two quotations by Poincaré and Hilbert, respectively ([13], p. 1003):

Metaparadox 2: (a) “Later generations will regard *Mengenlehre* [set theory] as a disease from which one has recovered.”

(b) “No one shall expel us from the paradise which Cantor created for us.”

PARADOXES INVOLVING CURVES. The notion of curve is, of course, fundamental in geometry. To Euclid it meant “breadthless length”. The collection of curves known to his contemporaries was small—the conic sections, the conchoid, the cissoid, the spiral, the quadratrix, and a very few others. The situation changed dramatically with the invention of analytic geometry in the 17th century. Now any equation in two variables came to represent a (plane) curve, although “seventeenth-century mathematicians did not have a uniform definition of the concept of curve (nor apparently did they feel the need for such a definition)” ([3], p. 296). The study of curves was pursued vigorously for the next three centuries, attracting some of the best mathematicians who attacked it by geometric, analytic, algebraic, arithmetic, and topological means.

“Pathological” functions introduced in the second half of the 19th century raised questions about the nature of curves. For example, in what sense does a continuous nowhere-differentiable function represent a curve? Jordan responded (in 1887) with what came to be the first formal definition of a curve (other than perhaps Euclid’s). To him a curve was the path of a continuously moving point. More precisely, it was $\{(f(t), g(t)) | f, g: [0, 1] \rightarrow \mathbb{R} \text{ are continuous functions}\}$.⁴ In 1890 Peano gave his famous and astounding example of a “space-filling curve;” that is, he exhibited a continuous mapping of the unit interval onto a square (including its interior). But that, according to Jordan’s definition, made the square into a curve—a not very desirable state of affairs. “How was it possible that intuition could so deceive us?”, wondered Poincaré ([24], p. 123). Jordan’s definition was too broad and had to be modified.

But Jordan’s definition also turned out to be too narrow. For we would surely want the graph of $y = \sin \frac{1}{x}$ and its limit points on the y -axis (i.e., $\{(x, \sin \frac{1}{x}) : x \in (-\infty, 0) \cup (0, \infty)\} \cup \{(0, y) : -1 \leq y \leq 1\}$) to be called a curve, but it is not the image of a continuously moving point.⁵

Metaparadox: How can a definition be both too broad and too narrow?

A satisfactory resolution of the dilemma was achieved (by Menger and Uryson) only in the 1920s. First one had to clarify the notion of dimension ([18]).⁶ When this was done, a curve was defined as a one-dimensional continuum (see [28]).⁷ The definition proved adequate until the 1970s when Mandelbrot introduced curves—his fractals—whose dimensions are fractions. See [9].

⁴It was in this context that he stated, and proved (incorrectly, as it later turned out) the celebrated “Jordan-Curve Theorem”.

⁵This is intuitively clear, although to prove it we need topological notions. See [11], p. 1968.

⁶That notion, too, was challenged by the paradoxical Peano curve which implied that a square is one-dimensional since it is the continuous image of the unit interval. Cantor’s proof of the one-one correspondence between an interval and a square also put to question the intuitive notion of dimension.

⁷A continuum is a closed, connected set of points.

PARADOXES INVOLVING DECOMPOSITION OF GEOMETRIC OBJECTS.

(a) In 1924 Banach and Tarski proved that a pea and the sun are equidecomposable. That is, the pea may be cut up into finitely many pieces⁸ which can be rearranged to yield the sun (in volume if not in substance). This is the celebrated *Banach-Tarski paradox* (see [27]). In fact, Banach and Tarski have shown that *any* two bounded sets in Euclidean space \mathbb{R}^n are equidecomposable if they contain interior points and if $n > 2$ (see [2], p. 351).⁹

Of course, the pieces into which the pea is cut in the Banach-Tarski decomposition are *not measurable*; that is, they do not have a volume. They are not the kinds of pieces that can be obtained using scissors or other cutting devices. They are obtained using the axiom of choice.

Metaparadox: How can simple assumptions (e.g. the axiom of choice) have such formidable consequences (e.g. the Banach-Tarski paradox)?

Of course, the axiom of choice may not be such a simple assumption after all (see [19]). But it would have been very helpful to the Delians of Greek antiquity ([27], p. v):

Delians: "How can we be rid of the plague?"

Delphic Oracle: "Construct a cubic altar double the size of the existing altar."

Banach and Tarski: "Can we use the axiom of choice?"

(b) At long last, the circle has been squared. This is no hoax. It is the title of an article which appeared recently in the reputable *Notices of the American Mathematical Society* ([10]). In 1988 the Hungarian mathematician Laczkovich showed that the circle can be decomposed into finitely many pieces which can be reassembled to give a square of equal area. But the pieces are not measurable (none has an area) and the decomposition is secured using the axiom of choice. See [10].

CONCLUDING REMARKS. We have presented a variety of mathematical paradoxes from different historical periods. They resulted from (among other things) debates and controversies among mathematicians, counterexamples to what were thought to be immutable notions, failures to see the need for tightening (broadening) a concept or broadening (tightening) a result, and the application of a "principle of continuity" which suggested the transferability of procedures from a given case to what appeared to be like cases. We saw that such paradoxical phenomena have had a very substantial impact on the development of mathematics through the refinement and reshaping of concepts, the broadening of existing theories and the rise of new ones. Moreover, this process is ongoing.

We have also suggested roles for paradoxes in the teaching and learning of mathematics. They can generate curiosity, increase motivation, create an effective environment for debate, encourage the examination of underlying assumptions, and show that faulty logic and erroneous arguments are not an uncommon feature of the mathematical enterprise.

⁸It was shown in the 1940s that five pieces suffice; in fact, no number less than five will do.

⁹If one allows for *denumerable* decompositions, then this result holds also for $n = 2$ (see [2], p. 351).

1. E. T. Bell, *The development of mathematics*, 2nd ed., McGraw-Hill, 1945.
2. L. M. Blumenthal, "A paradox, a paradox, a most ingenious paradox," *Amer. Math. Monthly* 47 (1940), 346–353.
3. H. J. M. Bos, "On the representation of curves in Descartes' *Géométrie*", *Arch. Hist. Ex. Sc.* 24 (1981), 295–338.
4. U. Bottazzini, *The higher calculus: a history of real and complex analysis from Euler to Weierstrass*, Springer-Verlag, 1986.
5. F. Cajori, "History of exponential and logarithmic concepts," *Amer. Math. Monthly* 20 (1913), several issues.
6. P. J. Davis, "Number," *Sc. Amer.* 211 (Sept. 1964), 51–59.
7. ———, *The mathematics of matrices*, Blaisdell, 1965.
8. C. H. Edwards, *The historical development of the calculus*, Springer-Verlag, 1979.
9. M. Gardner, "Mathematical games, in which 'monster' curves force redefinition of the word 'curve'," *Sc. Amer.* 235 (Dec. 1976), 124–133.
10. R. J. Gardner and S. Wagon, "At long last the circle has been squared", *Notic. Amer. Math. Soc.* 36 (1989), 1338–1343.
11. H. Hahn, "The crisis in intuition". In: *The world of mathematics*, ed. by J. R. Newman, Simon & Schuster, 1956, Vol. 3, pp. 1956–1976.
12. E. Kasner and J. R. Newman, *Mathematics and the imagination*, Simon & Schuster, 1967.
13. M. Kline, *Mathematical thought from ancient to modern times*, Oxford University Press, 1972.
14. Leapfrogs, *Imaginary logarithms*, E. G. M. Mann & Son (England), 1978.
15. ———, *Complex numbers*, E. G. M. Mann & Son (England), 1980.
16. J. Lützen, "Euler's vision of a generalized partial differential calculus for a generalized kind of function", *Math. Mag.* 56 (1983), 299–306.
17. P. Marchi, "The controversy between Leibniz and Bernoulli on the nature of the logarithms of negative numbers". In: *Akten des II Inter. Leibniz-Kongress* (Hanover, 1972), Bnd II, 1974, pp. 67–75.
18. K. Menger, "What is dimension?," *Amer. Math. Monthly* 50 (1943), 2–7.
19. G. H. Moore, *Zermelo's axiom of choice: its origins, development, and influence*, Springer-Verlag, 1982.
20. E. Nagel, "'Impossible numbers': a chapter in the history of modern logic," *Stud. in the Hist. of Ideas* 3 (1935), 429–474.
21. R. Remmert, *Theory of complex functions*, Springer-Verlag, 1991.
22. L. A. Steen, "New models of the real-number line," *Sc. Amer.* 225 (Aug. 1971), 92–99.
23. B. L. Van der Waerden, *Science awakening I*, Scholar's Bookshelf, 1988 (orig. 1954).
24. N. Ya. Vilenkin, *Stories about sets*, Academic Press, 1968.
25. K. Volkert, "Die Geschichte der pathologischen Funktionen—Ein Beitrag zur Entstehung der mathematischen Methodologie", *Arch. Hist. Ex. Sc.* 37 (1987), 193–232.
26. ———, "Zur Differenzierbarkeit stetiger Funktionen—Ampère's Beweis und seine Folgen", *Arch. Hist. Ex. Sc.* 40 (1989), 37–112.
27. S. Wagon, *The Banach-Tarski paradox*, Cambridge University Press, 1985.
28. G. T. Whyburn, "What is a curve?," *Amer. Math. Monthly* 49 (1942), 493–497.

Department of Mathematics & Statistics
 York University
 4700 Keele Street
 North York, Ontario
 CANADA M3J 1P3
 kleiner@vm1.yorku.ca

Dept. of Education in Science & Tech.
 Technion
 Haifa 32000, ISRAEL
 nitsa@techunix.technion.ac.il

Regions in the Complex Plane Containing the Eigenvalues of a Matrix

Richard A. Brualdi and Stephen Mellendorf

The Geršgorin circle theorem gives a region in the complex plane which contains all the eigenvalues of a square complex matrix. It is one of those rare instances of a theorem which is elegant and useful and which has a short, elementary proof. It is surprising that it hasn't made its way into many introductory texts on linear algebra. One reason for this neglect may be the fact that many (even most) mathematicians still regard linear algebra as being only about algebra. But modern linear algebra is more than algebra. It's linear systems, matrix theory and analysis, geometry, applications, numerics and, of course, algebra. The first course in linear algebra at most colleges and universities is to a great extent a service course for future scientists.

Geršgorin's theorem is not an algebraic theorem. It is a simple analytic theorem involving elementary inequalities derived from the basic algebraic eigenvalue/eigenvector equation. If $A = [a_{ij}]$ is a complex matrix of order n and λ is an eigenvalue of A , then there exists a nonzero vector $x = (x_1, x_2, \dots, x_n)^T$ in \mathbb{C}^n such that

$$Ax = \lambda x. \quad (1)$$

One of the components of x , say x_k , has the largest magnitude:

$$|x_k| = \max\{|x_i| : 1 \leq i \leq n\} > 0. \quad (2)$$

Choosing, the k th equation from among the n linear equations making up (1), we get

$$a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n = \lambda x_k$$

which, after rewriting, becomes

$$\sum_{j \neq k} a_{kj}x_j = (\lambda - a_{kk})x_k.$$

(Here, as later, the summation $\sum_{j \neq k}$ means the summation over all integers j between 1 and n excluding k .) Taking absolute values, and using the triangle inequality and (2), we then obtain

$$\begin{aligned} |\lambda - a_{kk}| |x_k| &= \left| \sum_{j \neq k} a_{kj}x_j \right| \\ &\leq \sum_{j \neq k} |a_{kj}| |x_j| \\ &\leq \left(\sum_{j \neq k} |a_{kj}| \right) |x_k|. \end{aligned}$$

Hence

$$|\lambda - a_{kk}| \leq \sum_{j \neq k} |a_{kj}|.$$

Defining

$$R_i = \sum_{j \neq i} |a_{ij}| = |a_{i1}| + \cdots + |a_{i,i-1}| + |a_{i,i+1}| + \cdots + |a_{in}| \quad (i = 1, 2, \dots, n),$$

the sum of the magnitudes of the entries of A in row i other than the entry on the main diagonal, we have proved the following theorem of Geršgorin.

Theorem 1. *Each eigenvalue of the matrix A of order n is in at least one of the disks*

$$\mathcal{D}_i(A) = \{z : |z - a_{ii}| \leq R_i\} \quad (1 \leq i \leq n) \quad (3)$$

in the complex plane. Equivalently, the n eigenvalues of A are contained in the region in the complex plane determined by

$$\mathcal{D}(A) = \bigcup_{i=1}^n \mathcal{D}_i(A). \quad (4)$$

We call (4) the *Geršgorin row-region* of A . By applying Theorem 1 to the transpose A^T of A we also obtain a *Geršgorin column-region* that contains all the eigenvalues of A . Let

$$S_j = \sum_{i \neq j} |a_{ij}| = |a_{1j}| + \cdots + |a_{j-1,j}| + |a_{j+1,j}| + \cdots + |a_{n,j}| \quad (1 \leq j \leq n),$$

the sum of the magnitudes of the entries of A in column j other than the entry on the main diagonal. Then each eigenvalue of A also lies in

$$\mathcal{D}'(A) = \bigcup_{j=1}^n \mathcal{D}'_j(A)$$

where

$$\mathcal{D}'_j(A) = \{z : |z - a_{jj}| \leq S_j\} \quad (1 \leq j \leq n). \quad (5)$$

Example. Let

$$A = \begin{bmatrix} 4 - 3i & i & 2 & -2 \\ i & -1 + i & 0 & 0 \\ 1 + i & -i & 5 + 6i & 2i \\ 1 & -2i & 2i & -5 - 5i \end{bmatrix}. \quad (6)$$

Then

$$R_1 = 5, R_2 = 1, R_3 = 3 + \sqrt{2}, \text{ and } R_4 = 5.$$

By Theorem 1, the eigenvalues of A are contained in the region $\mathcal{D}(A)$ shown in Figure 1.¹ We also have

$$S_1 = 2 + \sqrt{2}, \quad S_2 = 4, \quad S_3 = 4, \text{ and } S_4 = 4.$$

Hence the region $\mathcal{D}'(A)$ shown in Figure 2 also contains all the eigenvalues of A (so the eigenvalues of A all lie in $\mathcal{D}(A) \cap \mathcal{D}'(A)$). In fact, we computed² the

¹All figures were generated using Maple.

²The computation was done using MATLAB.

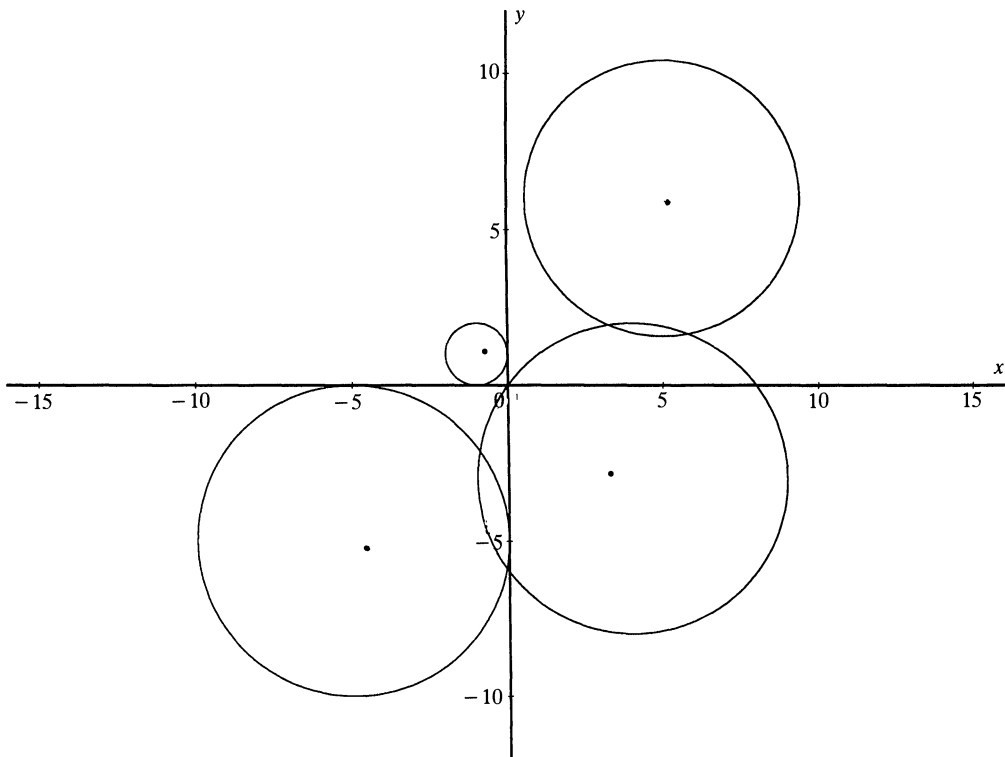


Figure 1

eigenvalues of A to be:

$$5.0353 + 5.9539i, 3.2940 - 2.8181i, -0.7229 + 1.0948i, -4.6065 - 5.2306i.$$

The eigenvalues correspond to the four points in Figures 1 and 2. The region $\mathcal{D}(A)$ has two connected components \mathcal{C}_1 and \mathcal{C}_2 where \mathcal{C}_1 is one disk and \mathcal{C}_2 is the union of three disks. It is a consequence of the fact that the eigenvalues of a matrix depend continuously on its entries that \mathcal{C}_1 contains exactly one eigenvalue of A and \mathcal{C}_2 contains the other three eigenvalues.

More generally, if A is a matrix of order n , and the Geršgorin row-region of A has two components \mathcal{C}_1 and \mathcal{C}_2 where \mathcal{C}_1 is the union of k disks and \mathcal{C}_2 is the union of the other $n - k$ disks, the \mathcal{C}_1 contains k of the eigenvalues of A and \mathcal{C}_2 contains the other $n - k$ eigenvalues of A . This can be seen by writing A in the form

$$A = D + B$$

where D is a diagonal matrix whose main diagonal is the same as the main diagonal as A and $B = A - D$. The Geršgorin row-regions of the matrices

$$A(\epsilon) = D + \epsilon B (0 \leq \epsilon \leq 1)$$

are all contained in the Geršgorin row-region of A , with the centers of the disks equal to the diagonal entries of A . Moreover, \mathcal{C}_1 contains k disks of $A(\epsilon)$ and \mathcal{C}_2 contains the other $n - k$ disks. We have $A(0) = D$ for which the disks degenerate

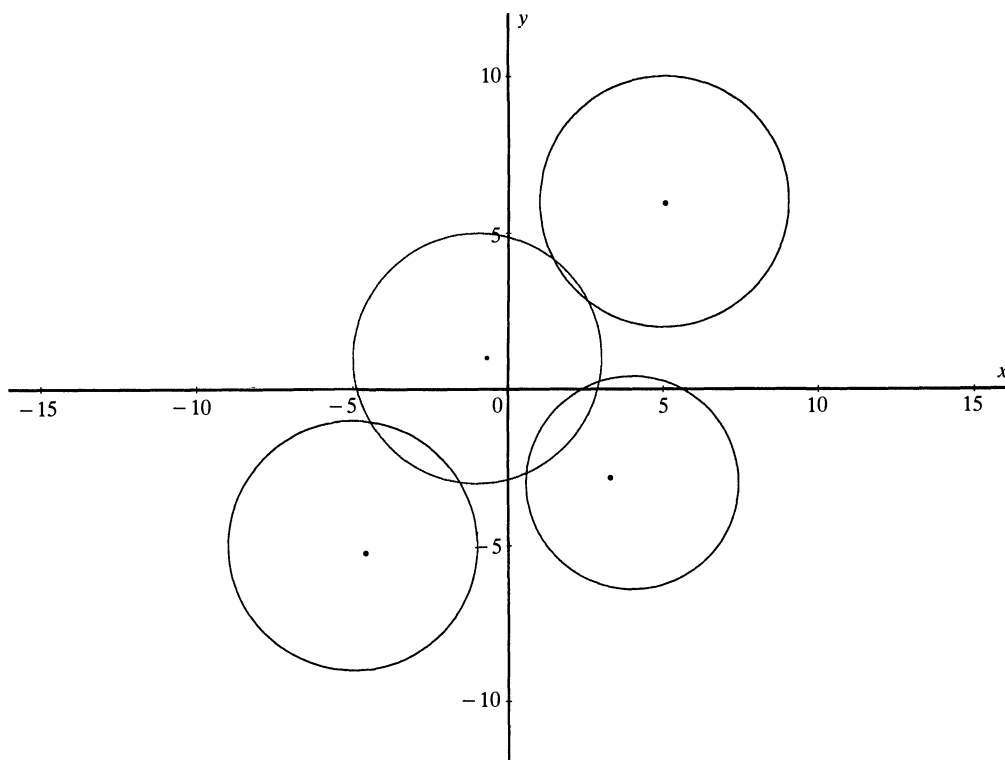


Figure 2

to the n points corresponding to the eigenvalues of D (the n diagonal entries of A). As ϵ increase from 0 to 1, the continuity of the eigenvalues implies that \mathcal{C}_1 always contains exactly k eigenvalues of $A(\epsilon)$. Since $A(1) = A$, \mathcal{C}_1 contains exactly k eigenvalues of A and \mathcal{C}_2 contains the other $n - k$ eigenvalues. Note that continuity does not imply that each disk always contains at least one eigenvalue (unless the disk itself is a connected component of the Geršgorin region). For example, let

$$A = \begin{bmatrix} 5 & -1 \\ 6 & 0 \end{bmatrix}.$$

Then the eigenvalues of A are 2 and 3, and neither of these is in the disk of radius 1 with center 5.

The matrix A of order n is *row-diagonally dominant* provided that

$$|a_{ii}| > R_i \quad (i = 1, 2, \dots, n), \quad (7)$$

that is, the distances from the centers of the Geršgorin row-disks to the origin are all greater than their radii. Thus if A is row-diagonally dominant, the Geršgorin row-region does not contain the origin and hence by Theorem 1, 0 is not an eigenvalue of A . Therefore if a matrix A is row-diagonally dominant, then A is invertible. Conversely, the fact that a row-diagonal dominant matrix A is invertible implies that the eigenvalues of A are contained in its Geršgorin row-region: If λ is an eigenvalue of A , then $\lambda I_n - A$ is not invertible and hence cannot be row-

diagonally dominant. Thus for some i we have

$$|\lambda - a_{ii}| \leq \sum_{k \neq i} |\lambda - a_{ik}| = \sum_{k \neq i} |a_{ik}| = R_i.$$

This is an instance of a correspondence that prevails between theorems implying that a matrix is invertible and theorems giving a region in the complex plane containing all the eigenvalues of a matrix. Similarly, if A is *column-diagonally dominant*, that is,

$$|a_{jj}| > S_j \quad (j = 1, 2, \dots, n),$$

then A is invertible.

In the next theorem we first obtain a condition that implies invertibility and then use it to obtain a region in the complex plane which contains the eigenvalues of a matrix. We use the fact that a matrix of order n is not invertible if and only if there is a nonzero vector x such that $xA = 0$ (equivalently, 0 is an eigenvalue of A).

Suppose that $x = (x_1, x_2, \dots, x_n)$ is a nonzero vector such that $xA = 0$. Let x_k be the component of x with the largest magnitude. Since $((1/x_k)x)A = 0$, we may assume that $|x_k| = 1$. The vector

$$(|x_1|, \dots, |x_{k-1}|, |x_{k+1}|, \dots, |x_n|) \quad (8)$$

thus belongs to the $(n - 1)$ -dimensional unit cube

$$Q_{n-1} = \{(a_1, a_2, \dots, a_{n-1}) : 0 \leq a_i \leq 1, (1 \leq i \leq n - 1)\}.$$

The cube Q_{n-1} is a convex set whose extreme points (vertices) are the 2^{n-1} $(n - 1)$ -tuples of 0's and 1's.

Lemma 2. Let $A = [a_{ij}]$ be a matrix of order n which is not invertible and let $x = (x_1, x_2, \dots, x_n)$ be a nonzero vector such that $xA = 0$. Assume that $1 = |x_k| \geq |x_i|$ for each $i = 1, 2, \dots, n$. Then $(|x_1|, \dots, |x_{k-1}|, |x_{k+1}|, \dots, |x_n|)$ is a point in Q_{n-1} satisfying the inequality

$$\sum_{i \neq k} |a_{ik}| |x_i| \geq |a_{kk}|. \quad (9)$$

Proof: Since $xA = 0$ we have

$$\sum_{i \neq k} a_{ik} x_i = -a_{kk} x_k.$$

Taking absolute values and using the triangle inequality and the fact that $|x_k| = 1$, we get

$$|a_{kk}| = \left| \sum_{i \neq k} a_{ik} x_i \right| \leq \sum_{i \neq k} |a_{ik}| |x_i|. \quad \square$$

Lemma 3. Let $A = [a_{ij}]$ be a matrix of order n which is not invertible and let $x = (x_1, x_2, \dots, x_n)$ be a nonzero vector such that $xA = 0$. Assume that $1 = |x_k| \geq |x_i|$ for each $i = 1, 2, \dots, n$. Then

$$\sum_{i \neq k} (|a_{ii}| - R_i) |x_i| \leq R_k - |a_{kk}|. \quad (10)$$

Proof: Since $x\mathcal{A} = 0$ we have

$$\sum_{i=1}^n a_{ij}x_i = 0 \quad (j = 1, 2, \dots, n).$$

Using the triangle inequality, we get

$$|a_{jj}| |x_j| \leq \sum_{i \neq j} |a_{ij}| |x_i| \quad (j = 1, 2, \dots, n).$$

Adding these inequalities and interchanging the order of summation we get

$$\begin{aligned} \sum_{j=1}^n |a_{jj}| |x_j| &\leq \sum_{j=1}^n \sum_{i \neq j} |a_{ij}| |x_i| \\ &= \sum_{i=1}^n \sum_{j \neq i} |a_{ij}| |x_i| \\ &= \sum_{i=1}^n R_i |x_i|. \end{aligned}$$

Subtracting the last sum from the first and using the fact that $|x_k| = 1$, we obtain (10). \square

The equation

$$\sum_{i \neq k} |a_{ik}| y_i = |a_{kk}|,$$

corresponding to the inequality (9), is the equation of a hyperplane H_k which partitions real $(n - 1)$ -dimensional space \mathbf{R}^{n-1} into a closed half-space

$$H_k^+ = \left\{ (y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_n) : \sum_{i \neq k} |a_{ik}| y_i \geq |a_{kk}| \right\} \quad (11)$$

and an open half-space

$$H_k^- = \left\{ (y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_n) : \sum_{i \neq k} |a_{ik}| y_i < |a_{kk}| \right\}. \quad (12)$$

Similarly the equation

$$\sum_{i \neq k} (|a_{ii}| - R_i) y_i = R_k - |a_{kk}|,$$

corresponding to the inequality (10), is the equation of a hyperplane J_k which partitions \mathbf{R}^{n-1} into a closed half-space

$$J_k^- = \left\{ (y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_n) : \sum_{i \neq k} (|a_{ii}| - R_i) y_i \leq R_k - |a_{kk}| \right\} \quad (13)$$

and an open half-space

$$J_k^+ = \left\{ (y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_n) : \sum_{i \neq k} (|a_{ii}| - R_i) y_i > R_k - |a_{kk}| \right\}. \quad (14)$$

Thus by Lemmas 2 and 3, a vector x with $x \neq 0$ such that $x\mathcal{A} = 0$ and such that $1 = |x_k| \geq |x_i| (i = 1, 2, \dots, n)$ determines a point (8) in the unit cube Q_{n-1} which is in both of the closed half-spaces H_k^+ and J_k^- . Hence if for each $k = 1, 2, \dots, n$ we assume that there is no point in the cube Q_{n-1} which is in both H_k^+ and J_k^- , then \mathcal{A} is an invertible matrix. Thus we have the following result which is equivalent to Theorem 3 of Pupkov [3].

Theorem 4. Let A be a matrix of order n such that

$$H_k^+ \cap J_k^- \cap Q_{n-1} = \emptyset \quad (k = 1, 2, \dots, n).$$

Then A is an invertible matrix.

Our goal is now to impose conditions on a matrix A which guarantee that the half-spaces H_k^+ and J_k^- do not have any common points in Q_{n-1} for each $k = 1, 2, \dots, n$. Such conditions will ensure the invertibility of the matrix. If our conditions are such that they guarantee that none of the half-spaces H_k^+ contains a point of Q_{n-1} , then A is invertible. Similarly, if our conditions imply that none of the half-spaces J_k^- contains a point of Q_{n-1} , then A is invertible.

For instance, assume that A is column-diagonally dominant. Then for each point $(y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_n)$ in Q_{n-1} we have

$$\sum_{i \neq k} |a_{ik}| y_i \leq \sum_{i \neq k} |a_{ik}| = S_k < |a_{kk}|$$

implying $H_k^+ \cap Q_{n-1}$ is empty for each $k = 1, 2, \dots, n$. Hence, as already noted, column-diagonal dominance implies invertibility.

Now assume that A is row-diagonally dominant. Then for each point $(y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_n)$ in Q_{n-1} we have

$$\sum_{i \neq k} (|a_{ii}| - R_i) y_i \geq 0.$$

Since $R_k - |a_{kk}| < 0$, we conclude that $J_k^- \cap Q_{n-1}$ is empty for each $k = 1, 2, \dots, n$. Hence, as also already noted, row-diagonal dominance implies invertibility.

We now show that a weakened form of row-diagonal dominance can be combined with a weakened form of column-diagonal dominance to ensure that H_k^+ and J_k^- have no common point in Q_{n-1} and thus to ensure invertibility. This result was obtained by Solov'ev [4] in a somewhat more complicated way. Before stating his result, we remark that we use the convention (already used above) that for each $k = 1, 2, \dots, n$ the $n - 1$ coordinates of Q_{n-1} can be written as $(y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_n)$.

Theorem 5. Let $A = [a_{ij}]$ be a matrix of order n and let r be an integer with $1 \leq r \leq n$. Assume that A satisfies both of the following conditions:

- (i) For each $j = 1, 2, \dots, n$

$$|a_{jj}| > S_j^{(r-1)}$$

where $S_j^{(r-1)}$ is the sum of the magnitudes of the $r - 1$ largest off-diagonal elements in column j ;

- (ii) For each set of r rows of A , the sum of the magnitudes of the diagonal elements in those rows is strictly greater than the sum of the magnitudes of all the off-diagonal elements in those rows.

Then A is invertible.

If in Theorem 5 we have $r = 1$, then condition (i) is vacuous and (ii) asserts that A is row-diagonally dominant. If $r = n$, then condition (i) asserts that A is column-diagonally dominant and (ii) is implied by column-diagonal dominance. Hence Theorem 5 generalizes the fact that row-diagonal dominance or column-diagonal dominance implies invertibility.

We note that condition (i) is equivalent to the statement that the magnitude of each diagonal element is strictly greater than the sum of the magnitudes of any $r - 1$ or fewer of the off-diagonal elements in its column. Condition (ii) is equivalent to the assertion that the sum of the r smallest of the numbers $|a_{11}| - R_1, |a_{22}| - R_2, \dots, |a_{nn}| - R_n$ is positive. Hence if (ii) is satisfied, at most $r - 1$ of these numbers are negative and hence any sum of $s \geq r$ of them is positive.

Proof of Theorem 5: Let k be an integer with $1 \leq k \leq n$ and consider the $(n - 1)$ -dimensional unit cube Q_{n-1} with coordinates written as $(y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_n)$. The hyperplane determined by the equation

$$y_1 + \dots + y_{k-1} + y_{k+1} + \dots + y_n = r - 1$$

determines two closed convex sets

$$Q_{n-1}^+ = \{(y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_n) \in Q_{n-1} : y_1 + \dots + y_{k-1} + y_{k+1} + \dots + y_n \geq r - 1\}$$

and

$$Q_{n-1}^- = \{(y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_n) \in Q_{n-1} : y_1 + \dots + y_{k-1} + y_{k+1} + \dots + y_n \leq r - 1\}.$$

The extreme points of Q_{n-1}^+ are the $(n - 1)$ -tuples of 0's and 1's with at least $r - 1$ 1's, while the extreme points of Q_{n-1}^- are the $(n - 1)$ -tuples of 0's and 1's with at most $r - 1$ 1's. We show that

- (a) Q_{n-1}^- is contained in H_k^- (and hence H_k^+ has no points in common with Q_{n-1}^-),

and that

- (b) Q_{n-1}^+ is contained in J_k^+ (and hence J_k^- has no points in common with Q_{n-1}^+).

This implies that there is no point in Q_{n-1} which is in $H_k^+ \cap J_k^-$. By Theorem 4 this will imply that A is invertible.

By condition (i) we know that for each subset I of $\{1, \dots, k - 1, k + 1, \dots, n\}$ of cardinality at most $r - 1$, we have

$$\sum_{i \in I} |a_{ik}| < |a_{kk}|.$$

Since the extreme points of Q_{n-1}^- are all the $(n - 1)$ -tuples of 0's and 1's with at most $r - 1$ 1's, it follows that

$$\sum_{i \neq k} |a_{ik}| y_i < |a_{kk}|$$

holds for all extreme points $(y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_n)$ of Q_{n-1}^- and hence for all points of Q_{n-1}^- . Therefore (a) holds.

By condition (ii) we know that for each subset L of $\{1, \dots, k - 1, k + 1, \dots, n\}$ of cardinality at least $r - 1$, we have

$$\sum_{i \in L \cup \{k\}} (|a_{ii}| - R_i) > 0,$$

and hence

$$\sum_{i \in L} (|a_{ii}| - R_i) > R_k - |a_{kk}|.$$

Since the extreme points of Q_{n-1}^+ are all the $(n-1)$ -tuples of 0's and 1's with at least $r-1$ 1's, we conclude that

$$\sum_{i \neq k} (|a_{ii}| - R_i) y_i > R_k - |a_{kk}|$$

holds for all extreme points $(y_1, \dots, y_{k-1}, y_{k+1}, \dots, y_n)$ of Q_{n-1}^+ and hence for all points of Q_{n-1}^+ . Therefore (b) holds. \square

Let λ be an eigenvalue of a matrix A . Then $\lambda I_n - A$ is not invertible. By applying Theorem 5 to the matrix $\lambda I_n - A$, we obtain another region in the complex plane which contains the eigenvalues of a matrix [4].

Theorem 6. *Let $A = [a_{ij}]$ be a matrix of order n and let r be an integer with $1 \leq r \leq n$. Then each eigenvalue of A is either in one of the disks*

$$\{z : |z - a_{jj}| \leq S_j^{(r-1)}\} \quad (j = 1, 2, \dots, n), \quad (15)$$

or in one of the regions

$$\left\{z : \sum_{i \in P} |z - a_{ii}| \leq \sum_{i \in P} R_i\right\} \quad (P \subseteq \{1, 2, \dots, n\}, |P| = r). \quad (16)$$

The region Γ_1 determined by the union of the disks in (15) is contained in the Geršgorin column-region of A . There are $\binom{n}{r}$ regions in (16) and each region is an 'average' of r of the disks making up the Geršgorin row-region. Let

$$\left\{z : \sum_{i \in P} |z - a_{ii}| \leq \sum_{i \in P} R_i\right\} \quad (17)$$

be one of the regions in (16). If μ is a complex number such that $|\mu - a_{ii}| > R_i$ for each i in P , then

$$\sum_{i \in P} |\mu - a_{ii}| > \sum_{i \in P} R_i$$

and hence μ is not contained in (17). It follows that the region in (17) is contained in the union of the r disks with centers at a_{ii} and radii equal to R_i with $i \in P$. In particular this implies that the union Γ_2 of the regions defined by (16) is contained in the Geršgorin row-region of A . Thus Theorem 6 determines two regions Γ_1 and Γ_2 whose union $\Gamma_1 \cup \Gamma_2$ contains all of the eigenvalues of A where Γ_1 is contained in the Geršgorin column-region of A and Γ_2 is contained in the Geršgorin row-region.

The regions in (16) are in general very hard to describe. But in the case $r = 2$ we obtain the regions

$$\{z : |z - a_{ii}| + |z - a_{jj}| \leq R_i + R_j\} \quad (1 \leq i < j \leq n). \quad (18)$$

If the two disks

$$|z - a_{ii}| \leq R_i \quad \text{and} \quad |z - a_{jj}| \leq R_j$$

are disjoint, then the region (18) is empty. Otherwise, the region (18) is bounded by an ellipse with foci at a_{ii} and a_{jj} and with major axis $R_i + R_j$ and minor axis

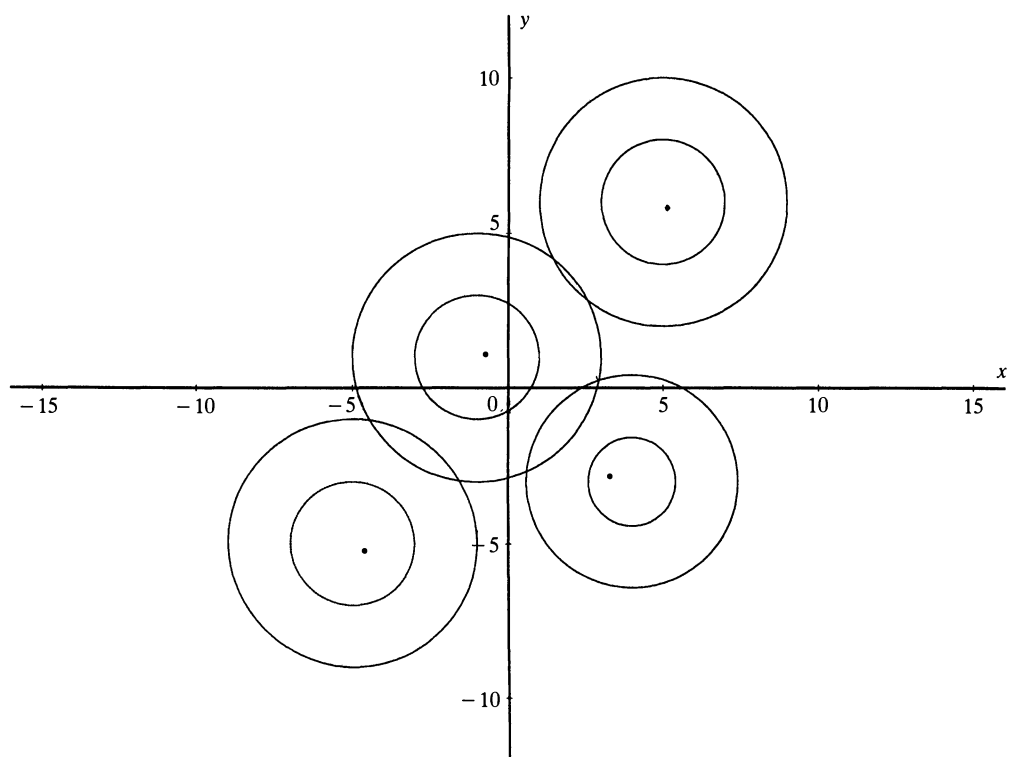


Figure 3

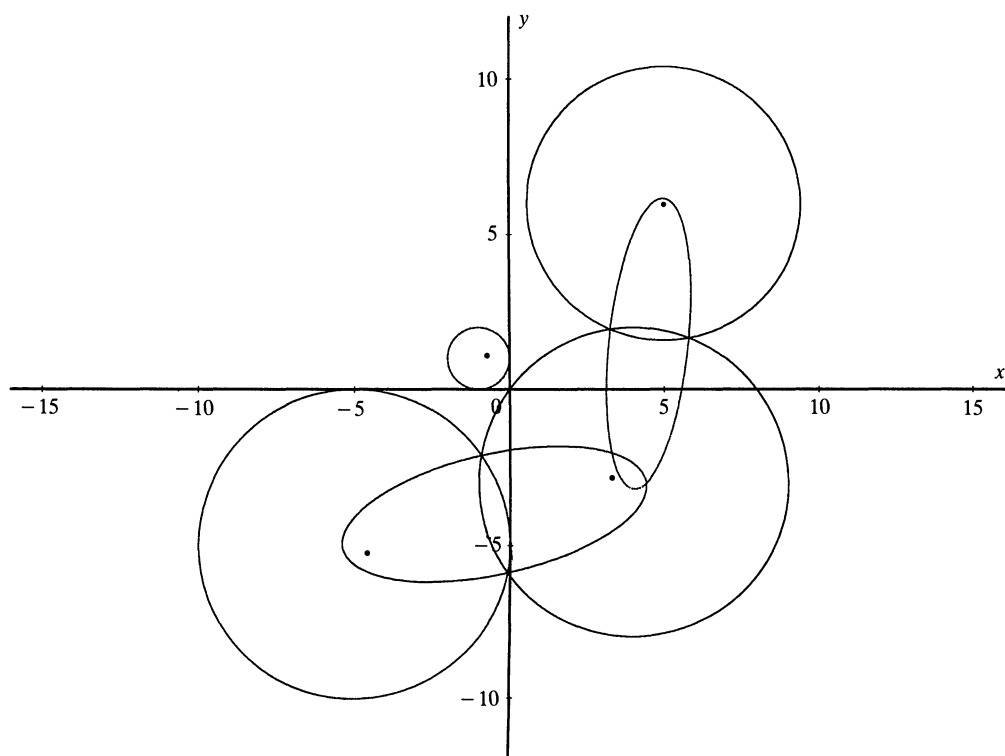


Figure 4

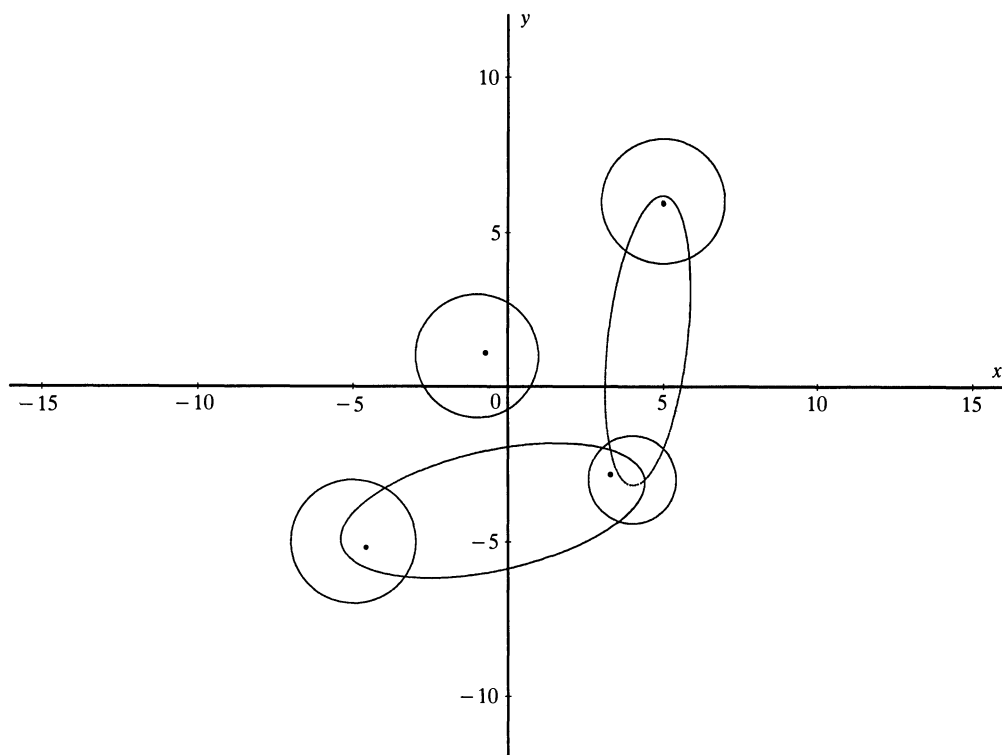


Figure 5

$\sqrt{(R_i + R_j)^2 - |a_{ii} - a_{jj}|^2}$. This ellipse passes through the two points of intersection of the circles bounding the two disks. In the case that one disk is contained in the other, then the ellipse contains the smaller disk and is contained in the larger.

Let A be the matrix in (6) and let $r = 2$. In Figure 3 we show the region Γ_1 as a subset of the Geršgorin column-region of Figure 2, and in Figure 4 we show the region Γ_2 as a subset of the Geršgorin row-region of Figure 1. In Figure 5 we show the region $\Gamma_1 \cup \Gamma_2$ which by Theorem 6 contains the eigenvalues of A .

Other generalizations of Geršgorin's theorem, and many references, can be found in [1] and [2].

REFERENCES

1. R. A. Brualdi, Matrices, eigenvalues, and directed graphs, *Linear and Multilin. Alg.*, 11 (1982), 143–165.
2. R. A. Brualdi and H. J. Ryser, *Combinatorial Matrix Theory*, Cambridge University Press, New York, 1991.
3. V. A. Pupkov, Some sufficient conditions for the non-degeneracy of matrices, *U.S.S.R. Comput. Math. and Math. Phys.*, 24 (1984), 86–89.
4. V. A. Solov'ev, A generalization of Geršgorin's theorem, *Math. USSR Izvestiya*, 23 (1984), 545–559.

Department of Mathematics
University of Wisconsin
Madison, WI 53706
brualdi@math.wisc.edu
mellendo@math.wisc.edu

Characterization of Solvable Quintics

$$x^5 + ax + b$$

Blair K. Spearman and Kenneth S. Williams

We consider the quintic equation

$$x^5 + ax + b = 0, \quad (1)$$

where a and b are nonzero rational numbers. In general the roots of (1) cannot be expressed as algebraic functions of the coefficients a and b . We will characterize completely those irreducible quintics $x^5 + ax + b$ which are solvable by radicals. We do this by extending Cardano's familiar method of solving the cubic equation $x^3 + ax + b = 0$. We begin by recalling Cardano's method in a way which enables us to apply it to the quintic equation (1).

If u_1, u_2 are complex numbers and ω is a complex cube root of unity, expanding the product

$$(x - (u_1 + u_2))(x - (\omega u_1 + \omega^2 u_2))(x - (\omega^2 u_1 + \omega u_2)), \quad (2)$$

we obtain the polynomial

$$x^3 - 3u_1 u_2 x - (u_1^3 + u_2^3). \quad (3)$$

As $x_j = \omega^j u_1 + \omega^{2j} u_2$ ($j = 0, 1, 2$) is a root of the cubic polynomial (2), substituting it into (3), we obtain the identity valid for $j = 0, 1, 2$

$$(\omega^j u_1 + \omega^{2j} u_2)^3 - 3u_1 u_2 (\omega^j u_1 + \omega^{2j} u_2) - (u_1^3 + u_2^3) = 0.$$

Thus the cubic $x^3 + ax + b = 0$ has the three solutions $x_j = \omega^j u_1 + \omega^{2j} u_2$ ($j = 0, 1, 2$), where u_1^3 and u_2^3 are determined from $u_1^3 + u_2^3 = -b$, $u_1^3 u_2^3 = -(a/3)^3$.

An obvious generalization of this is to consider the quintic polynomial

$$\prod_{j=0}^4 (x - (\omega^j u_1 + \omega^{4j} u_2)), \quad (4)$$

where ω is now a complex fifth root of unity. Expanding the product (4), and proceeding as above, we find that the quintic $x^5 + ax^3 + (a^2/5)x + b$ (sometimes called DeMoivre's quintic) has the solutions $x_j = \omega^j u_1 + \omega^{4j} u_2$, $j = 0, 1, 2, 3, 4$, where u_1^5 and u_2^5 are determined from $u_1^5 + u_2^5 = -b$, $u_1^5 u_2^5 = -(a/5)^5$.

We refine this method by considering instead of (4) the quintic polynomial

$$\prod_{j=0}^4 (x - (\omega^j u_1 + \omega^{2j} u_2 + \omega^{3j} u_3 + \omega^{4j} u_4)), \quad (5)$$

where u_1, u_2, u_3, u_4 are nonzero real numbers and ω is a complex fifth root of unity. Multiplying out (5) is somewhat more challenging than (4), so MAPLE was employed to do the work. Replacing x by $\omega^j u_1 + \omega^{2j} u_2 + \omega^{3j} u_3 + \omega^{4j} u_4$ in the

expanded product, we obtain the identity valid for $j = 0, 1, 2, 3, 4$

$$\begin{aligned}
 & (\omega^j u_1 + \omega^{2j} u_2 + \omega^{3j} u_3 + \omega^{4j} u_4)^5 \\
 & - 5U(\omega^j u_1 + \omega^{2j} u_2 + \omega^{3j} u_3 + \omega^{4j} u_4)^3 \\
 & - 5V(\omega^j u_1 + \omega^{2j} u_2 + \omega^{3j} u_3 + \omega^{4j} u_4)^2 \\
 & + 5W(\omega^j u_1 + \omega^{2j} u_2 + \omega^{3j} u_3 + \omega^{4j} u_4) \\
 & + 5(X - Y) - Z \\
 & = 0,
 \end{aligned} \tag{6}$$

where

$$U = u_1 u_4 + u_2 u_3,$$

$$V = u_1 u_2^2 + u_2 u_4^2 + u_3 u_1^2 + u_4 u_3^2,$$

$$W = u_1^2 u_4^2 + u_2^2 u_3^2 - u_1^3 u_2 - u_2^3 u_4 - u_3^3 u_1 - u_4^3 u_3 - u_1 u_2 u_3 u_4,$$

$$X = u_1^3 u_3 u_4 + u_2^3 u_1 u_3 + u_3^3 u_2 u_4 + u_4^3 u_1 u_2,$$

$$Y = u_1 u_3^2 u_4^2 + u_2 u_1^2 u_3^2 + u_3 u_2^2 u_4^2 + u_4 u_1^2 u_2^2,$$

$$Z = u_1^5 + u_2^5 + u_3^5 + u_4^5.$$

The essential ingredient of the proof of our characterization of solvable quintic trinomials is the determination of real algebraic numbers u_1, u_2, u_3, u_4 satisfying

$$u_1 u_4 + u_2 u_3 = 0, \tag{7}$$

$$u_1 u_2^2 + u_2 u_4^2 + u_3 u_1^2 + u_4 u_3^2 = 0, \tag{8}$$

$$5(u_1^2 u_4^2 + u_2^2 u_3^2 - u_1^3 u_2 - u_2^3 u_4 - u_3^3 u_1 - u_4^3 u_3 - u_1 u_2 u_3 u_4) = a, \tag{9}$$

and

$$\begin{aligned}
 & 5((u_1^3 u_3 u_4 + u_2^3 u_1 u_3 + u_3^3 u_2 u_4 + u_4^3 u_1 u_2) \\
 & - (u_1 u_3^2 u_4^2 + u_2 u_1^2 u_3^2 + u_3 u_2^2 u_4^2 + u_4 u_1^2 u_2^2)) \\
 & - (u_1^5 + u_2^5 + u_3^5 + u_4^5) = b,
 \end{aligned} \tag{10}$$

so that the quintic polynomial (5) becomes $x^5 + ax + b$ and has the roots

$$x_j = (\omega^j u_1 + \omega^{2j} u_2 + \omega^{3j} u_3 + \omega^{4j} u_4) \quad (j = 0, 1, 2, 3, 4). \tag{11}$$

Theorem. *Let a and b be rational numbers such that the quintic trinomial $x^5 + ax + b$ is irreducible. Then the equation $x^5 + ax + b = 0$ is solvable by radicals if and only if there exist rational numbers $\epsilon (= \pm 1)$, $c (\geq 0)$ and $e (\neq 0)$ such that*

$$a = \frac{5e^4(3 - 4\epsilon c)}{c^2 + 1}, \quad b = \frac{-4e^5(11\epsilon + 2c)}{c^2 + 1}, \tag{12}$$

in which case the roots of $x^5 + ax + b = 0$ are

$$x_j = e(\omega^j u_1 + \omega^{2j} u_2 + \omega^{3j} u_3 + \omega^{4j} u_4) \quad (j = 0, 1, 2, 3, 4), \tag{13}$$

where $\omega = \exp(2\pi i/5)$ and

$$u_1 = \left(\frac{v_1^2 v_3}{D^2} \right)^{1/5}, \quad u_2 = \left(\frac{v_3^2 v_4}{D^2} \right)^{1/5}, \quad u_3 = \left(\frac{v_2^2 v_1}{D^2} \right)^{1/5}, \quad u_4 = \left(\frac{v_4^2 v_2}{D^2} \right)^{1/5}, \quad (14)$$

$$\begin{cases} v_1 = \sqrt{D} + \sqrt{D - \epsilon\sqrt{D}}, & v_2 = -\sqrt{D} - \sqrt{D + \epsilon\sqrt{D}}, \\ v_3 = -\sqrt{D} + \sqrt{D + \epsilon\sqrt{D}}, & v_4 = \sqrt{D} - \sqrt{D - \epsilon\sqrt{D}}, \end{cases} \quad (15)$$

$$D = c^2 + 1. \quad (16)$$

Proof: We begin by supposing that the irreducible quintic polynomial $x^5 + ax + b$ is solvable by radicals. Thus the resolvent sextic of $x^5 + ax + b$, namely,

$$x^6 + 8ax^5 + 40a^2x^4 + 160a^3x^3 + 400a^4x^2 + (512a^5 - 3125b^4)x + (256a^6 - 9375ab^4)$$

has a rational root r [1, Theorem 1]. Hence r satisfies

$$(r + 2a)^4(r^2 + 16a^2) - 5^5b^4(r + 3a) = 0, \quad (17)$$

which shows that $r \neq -2a, -3a$ as $a \neq 0$. We define the nonnegative rational number c and the nonzero rational number e by

$$\epsilon c = \frac{3r - 16a}{4(r + 3a)}, \quad e = \frac{-5b\epsilon}{2(r + 2a)}, \quad \text{where } \epsilon = \pm 1. \quad (18)$$

Then

$$\begin{aligned} c^2 + 1 &= \frac{25(r^2 + 16a^2)}{16(r + 3a)^2}, \\ 3 - 4\epsilon c &= \frac{25a}{r + 3a}, \\ 11\epsilon + 2c &= \frac{25(r + 2a)\epsilon}{2(r + 3a)}, \end{aligned}$$

so that

$$\frac{5e^4(3 - 4\epsilon c)}{c^2 + 1} = \frac{5^5ab^4(r + 3a)}{(r + 2a)^4(r^2 + 16a^2)} = a,$$

and

$$\frac{-4e^5(11\epsilon + 2c)}{c^2 + 1} = \frac{5^5b^5(r + 3a)}{(r + 2a)^4(r^2 + 16a^2)} = b,$$

giving the required parametrization.

We now show that the irreducible quintic trinomial

$$x^5 + \frac{5e^4(3 - 4\epsilon c)}{c^2 + 1}x - \frac{4e^5(11\epsilon + 2c)}{c^2 + 1} \quad (19)$$

with $e = 1$ is solvable by radicals with roots given by (11). In fact it is not necessary to assume that the quintic is irreducible. For general e the transformation $x \rightarrow ex$

gives the required result (13). From (15) we see that

$$\begin{cases} v_1 + v_4 = 2\sqrt{D}, & v_2 + v_3 = -2\sqrt{D}, \\ v_1v_4 = \epsilon\sqrt{D}, & v_2v_3 = -\epsilon\sqrt{D}, \end{cases} \quad (20)$$

and so

$$\begin{cases} v_1 + v_2 + v_3 + v_4 = 0, \\ v_1v_4 + v_2v_3 = 0. \end{cases} \quad (21)$$

Further, from (14), we obtain

$$u_1^5 = \frac{v_1^2v_3}{D^2}, \quad u_2^5 = \frac{v_3^2v_4}{D^2}, \quad u_3^5 = \frac{v_2^2v_1}{D^2}, \quad u_4^5 = \frac{v_4^2v_2}{D^2}. \quad (22)$$

Easy calculations making use of (20) and (22) yield

$$u_1u_4 = -\frac{\epsilon}{\sqrt{D}}, \quad u_2u_3 = \frac{\epsilon}{\sqrt{D}}, \quad (23)$$

$$u_1u_2^2 = \frac{v_3}{D}, \quad u_3^2u_4 = \frac{v_2}{D}, \quad u_1^2u_3 = \frac{v_1}{D}, \quad u_4^2u_2 = \frac{v_4}{D}, \quad (24)$$

and

$$u_1^3u_2 = \frac{\epsilon v_1v_3}{D\sqrt{D}}, \quad u_2^3u_4 = -\frac{\epsilon v_3v_4}{D\sqrt{D}}, \quad u_3^3u_1 = -\frac{\epsilon v_2v_1}{D\sqrt{D}}, \quad u_4^3u_3 = \frac{\epsilon v_4v_2}{D\sqrt{D}}, \quad (25)$$

which give the required equations (7) and (8) in view of (21). From (15), (22), (23), (24) and (25), we deduce that

$$\begin{aligned} & 5(u_1^2u_4^2 + u_2^2u_3^2 - u_1^3u_2 - u_2^3u_4 - u_3^3u_1 - u_4^3u_3 - u_1u_2u_3u_4) \\ &= \frac{5(3 - 4\epsilon\sqrt{D} - 1)}{D} = \frac{5(3 - 4\epsilon c)}{c^2 + 1} \end{aligned} \quad (26)$$

and

$$\begin{aligned} & 5((u_1^3u_3u_4 + u_2^3u_1u_3 + u_3^3u_2u_4 + u_4^3u_1u_2) \\ & - (u_1u_3^2u_4^2 + u_2u_1^2u_3^2 + u_3u_2^2u_4^2 + u_4u_1^2u_2^2)) \\ & - (u_1^5 + u_2^5 + u_3^5 + u_4^5) = -\frac{(44\epsilon + 8\sqrt{D} - 1)}{D} = -\frac{4(11\epsilon + 2c)}{c^2 + 1}, \end{aligned} \quad (27)$$

which are the required equations (9) and (10). This proves that

$$x^5 + \frac{5(3 - 4\epsilon c)}{c^2 + 1}x - \frac{4(11\epsilon + 2c)}{c^2 + 1}$$

is solvable by radicals and has the roots given in (11).

The discriminant of the trinomial quintic $x^5 + ax + b$ is $4^4a^5 + 5^5b^4$ [2, p. 259]. The equation $x^5 + ax + b = 0$ has exactly one real root if $4^4a^5 + 5^5b^4 > 0$ [3, p. 113]. The discriminant of the quintic (19) is

$$\frac{4^45^5e^{20}}{D^5}(4\epsilon c^3 - 84c^2 - 37\epsilon c - 122)^2 > 0 \quad (28)$$

so that the quintic (19) has exactly one real root. Suppose now that (19) is

irreducible over \mathcal{Q} . By the Theorem, (19) is solvable by radicals, and so its Galois group is solvable. Hence its Galois group is isomorphic to the Frobenius group F_{20} of order 20, the dihedral group D_5 of order 10, or to the cyclic group of order 5. However (19) has complex roots, so its Galois group cannot be cyclic of order 5. By [1, Theorem 2] the Galois group of (19) is the dihedral group D_5 of order 10 if and only if $5D$ is a perfect square in \mathcal{Q} . Otherwise the Galois group is the Frobenius group F_{20} of order 20.

We close with five examples.

Example 1. We consider the quintic $f_1(x) = x^5 - 5x + 12$, which is irreducible as $f_1(x - 2)$ is 5-Eisenstein. The resolvent sextic of f_1 is

$$x^6 - 40x^5 + 1000x^4 + 20000x^3 + 250000x^2 - 66400000x + 976000000,$$

which has the rational root $r = 40$. From (18) we see that $\epsilon = 1$, $c = 2$, $e = -1$, so that by (16) $D = 5$. Since $5D = 5^2$ the Galois group of f_1 is D_5 . By the Theorem the unique real root of f_1 is

$$\begin{aligned} x = & - \left(\frac{(\sqrt{5} + \sqrt{5 - \sqrt{5}})^2 (-\sqrt{5} + \sqrt{5 + \sqrt{5}})}{25} \right)^{1/5} \\ & - \left(\frac{(-\sqrt{5} + \sqrt{5 + \sqrt{5}})^2 (\sqrt{5} - \sqrt{5 - \sqrt{5}})}{25} \right)^{1/5} \\ & - \left(\frac{(-\sqrt{5} - \sqrt{5 + \sqrt{5}})^2 (\sqrt{5} + \sqrt{5 - \sqrt{5}})}{25} \right)^{1/5} \\ & - \left(\frac{(\sqrt{5} - \sqrt{5 - \sqrt{5}})^2 (-\sqrt{5} - \sqrt{5 + \sqrt{5}})}{25} \right)^{1/5} \end{aligned}$$

A little manipulation shows that this root can be rewritten as

$$x = \frac{1}{5}(R_1^{1/5} + R_2^{1/5} + R_3^{1/5} + R_4^{1/5}),$$

where R_1, R_2, R_3, R_4 are given at the bottom of page 399 of [1].

Example 2. We take $f_2(x) = x^5 + 15x + 12$, which is irreducible as $f_2(x)$ is 3-Eisenstein. The resolvent sextic of f_2 is

$$(x + 30)^4(x^2 + 1800) - 2^8 \cdot 3^4 \cdot 5^4(x + 45),$$

which has the rational root $r = 0$. Hence, by (16) and (18), we have $\epsilon = -1$, $c = 4/3$, $e = 1$, $D = 25/9$. Since $5D$ is not the square of a rational number, the Galois group of f_2 is F_{20} . By the Theorem the unique real root of f_2 is

$$\begin{aligned} x = & \left(\frac{-75 - 21\sqrt{10}}{125} \right)^{1/5} + \left(\frac{225 - 72\sqrt{10}}{125} \right)^{1/5} \\ & + \left(\frac{225 + 72\sqrt{10}}{125} \right)^{1/5} + \left(\frac{-75 + 21\sqrt{10}}{125} \right)^{1/5} \end{aligned}$$

in agreement with the more complicated formula given at the top of page 399 in [1].

Example 3. Here we take $\epsilon = 1$, $e = 5/2$, $c = 7/24$, so $D = 1 + (\frac{7}{24})^2 = (\frac{25}{24})^2$, and the quintic (19) is $f_3(x) = x^5 + 330x - 4170$, which is irreducible as $f_3(x)$ is 5-Eisenstein. Since $5D = 5^5/(2^6 \cdot 3^2)$ the Galois group of f_3 is F_{20} . By the Theorem the unique real root of f_3 is

$$x = 54^{1/5} + 12^{1/5} + 648^{1/5} - 144^{1/5}.$$

Example 4. Here we take $\epsilon = -1$, $e = 1$, $c = 11/2$, so $D = 125/4$, and the quintic (19) is $f_4(x) = x^5 + 4x$, which is clearly reducible. However, by the remark preceeding (20), the roots of $x^5 + 4x = 0$, namely $x = 0, \pm(1 \pm i)$, are given by (13). Here

$$v_1 = \frac{1}{2}(5\sqrt{5} + \sqrt{5}\sqrt{25 + 2\sqrt{5}}), \quad v_3 = \frac{1}{2}(-5\sqrt{5} + \sqrt{5}\sqrt{25 - 2\sqrt{5}}),$$

$$\begin{aligned} \frac{v_1^2 v_3}{D^2} &= \frac{1}{5^5} (1000 - 500\sqrt{5} + 180\sqrt{25 + 2\sqrt{5}} - 240\sqrt{25 - 2\sqrt{5}}) \\ &= \frac{1}{5^5} (1000 - 500\sqrt{5} + 120\sqrt{5 + 2\sqrt{5}} - 660\sqrt{5 - 2\sqrt{5}}), \end{aligned}$$

and

$$u_1 = \left(\frac{v_1^2 v_3}{D^2} \right)^{1/5} = \frac{1}{5} (-\sqrt{5} - \sqrt{5 - 2\sqrt{5}}).$$

The conjugates of u_1 are

$$u_2 = \frac{1}{5} (\sqrt{5} - \sqrt{5 + 2\sqrt{5}}),$$

$$u_3 = \frac{1}{5} (\sqrt{5} + \sqrt{5 + 2\sqrt{5}}),$$

$$u_4 = \frac{1}{5} (-\sqrt{5} + \sqrt{5 - 2\sqrt{5}}).$$

Clearly $x_0 = u_1 + u_2 + u_3 + u_4 = 0$. Further, as

$$\omega = \exp(2\pi i/5) = ((\sqrt{5} - 1) + i\sqrt{10 + 2\sqrt{5}})/4,$$

we have

$$\begin{aligned} x_1 &= u_1\omega + u_2\omega^2 + u_3\omega^3 + u_4\omega^4 \\ &= \frac{1}{20} ((-x - y)(x - 1 + i(y + z)) + (x - z)(-x - 1 - i(y - z)) \\ &\quad + (x + z)(-x - 1 + i(y - z)) + (-x + y)(x - 1 - i(y + z))), \end{aligned}$$

where $x = \sqrt{5}$, $y = \sqrt{5 - 2\sqrt{5}}$, $z = \sqrt{5 + 2\sqrt{5}}$. Simplifying the expression for x_1 , we deduce

$$x_1 = \frac{1}{20} (-4x^2 - 2iy^2 - 2iz^2) = \frac{-20 - 20i}{20} = -1 - i.$$

We leave it to the reader to show that $x_2 = 1 + i$, $x_3 = 1 - i$, $x_4 = -1 + i$.

Example 5. Let p be a prime with $p \equiv 3 \pmod{4}$. We show using the Theorem that the quintic equation $x^5 + 2px + 2p^2 = 0$ is not solvable by radicals. We first observe that $x^5 + 2px + 2p^2$ is 2-Eisenstein so that it is irreducible. Suppose however that the equation is solvable by radicals. Then, by the Theorem, there

exist rational numbers $\epsilon (= \pm 1)$, $c (\geq 0)$ and $e (\neq 0)$ such that

$$2p = \frac{5e^4}{c^2 + 1}(3 - 4\epsilon c), \quad (29)$$

$$2p^2 = -\frac{4e^5}{c^2 + 1}(11\epsilon + 2c). \quad (30)$$

Expressing the rational numbers c and e in the form $c = m/n$ and $e = r/s$, where m, n, r, s are integers with $\gcd(m, n) = \gcd(r, s) = 1$, and appealing to (29) and (30), we obtain

$$2p(m^2 + n^2)s^4 = 5r^4(3n - 4\epsilon m)n, \quad (31)$$

$$2p^2(m^2 + n^2)s^5 = -4r^5(11n\epsilon + 2m)n. \quad (32)$$

As p is a prime $\equiv 3 \pmod{4}$ and $\gcd(m, n) = 1$, p does not divide $m^2 + n^2$. Further, as $\gcd(r, s) = 1$, it is clear from (31) that p does not divide r . Let $p^\alpha, p^\beta, p^\gamma, p^\delta$ be the exact powers of p dividing $n, 3n - 4\epsilon m, 11n\epsilon + 2m, s$ respectively. As p does not divide both of n and $3n - 4\epsilon m$ we see that α or $\beta = 0$. Similarly α or $\gamma = 0$ and β or $\gamma = 0$. Equating powers of p on both sides of (31) and (32), we obtain

$$\begin{cases} 1 + 4\delta = \alpha + \beta, \\ 2 + 5\delta = \alpha + \gamma, \end{cases}$$

which contradicts that at least two of α, β, γ are 0. Hence the equation $x^5 + 2px + 2p^2 = 0$ is not solvable by radicals.

Other examples of solvable quintics are given below together with their Galois groups.

$\epsilon = 1,$	$e = -1,$	$c = 2/11$	$x^5 + 11x + 44$	D_5
$\epsilon = 1,$	$e = -1,$	$c = 0$	$x^5 + 15x + 44$	F_{20}
$\epsilon = -1,$	$e = 1,$	$c = 1/2$	$x^5 + 20x + 32$	D_5
$\epsilon = 1,$	$e = -2,$	$c = 7$	$x^5 - 40x + 64$	F_{20}

REFERENCES

1. D. S. Dummit, Solving solvable quintics, *Math. Comp.* 57 (1991), 387–401.
2. N. Jacobson, *Basic Algebra 1*, W. H. Freeman and Company, San Francisco (1974).
3. J. V. Uspensky, *Theory of Equations*, McGraw-Hill Book Company, Inc., New York, Toronto, London (1948).

Blair K. Spearman
Department of Mathematics
Okanagan University College
Kelowna, B.C., Canada
V1Y 4X8

Kenneth S. Williams
Department of Mathematics and Statistics
Carleton University
Ottawa, Ontario, Canada
K1S 5B6

A Halmos Problem and a Related Problem

John B. Cosgrave

To my daughters Catherine and Marie for their 18th and 21st birthdays, and for my first year students Clare Connolly and Orla Walsh without whom there would have been no story to tell.

The November '92 issue of the *Monthly* has a review by Stan Wagon of Paul Halmos' 'Problems for Mathematicians Young and Old,' and one of those which he quotes is:

"Which positive integers are sums of three or more consecutive positive integers?", and then says: several of these problems have surprising answers and I won't spoil your fun too much, but to whet your appetite for the book here is the answer to the above problem: all integers except for the primes and the powers of 2 (surprising, and surprisingly easy to prove).

I first saw this problem in a British Sunday newspaper, the *Sunday Times*, about fifteen years ago—I don't know who first thought of it, perhaps some British reader could provide a reference—but only last October did I present it, for the first time to any of my students, to my first year students, who are training to be primary school teachers.

Then I was introducing them to Number Theory and had asked them to find some primes which are a power of 2 plus 1. They quickly came up with 3, 5 and 17 which are $(2^1 + 1)$, $(2^2 + 1)$ and $(2^4 + 1)$, and one of them said 'there's a pattern!' Good! Namely? 'The next one will be $(2^8 + 1)$.' They checked that it is prime, and I asked how could they be sure that it is the next prime of this type. They tested all the 'missing' ones, $(2^3 + 1)$, $(2^5 + 1)$, $(2^6 + 1)$ and $(2^7 + 1)$, all composite. [Later I showed, for example, $(2^{35} + 1)$ can be seen to be composite: $(2^{35} + 1) = (2^5 + 1)(2^{30} - 2^{25} + 2^{20} - 2^{15} + 2^{10} - 2^5 + 1)$].

And the next one? I was of course, given $(2^{16} + 1)$; and then? Well—I was told—the next one is $(2^{32} + 1)$. It was then I let out that they had just rediscovered Fermat's false claim with respect to his primes (that the 'Fermat numbers' $F_n = 2^{2^n} + 1$, are prime for all $n = 0, 1, 2, 3, \dots$) and got them to verify that 641 divides $(2^{32} + 1)$ (Euler). I also asked them to find primes that are a power of 2 less 1; here they quickly found 3, 7, 31 and 127 which are $(2^2 - 1)$, $(2^3 - 1)$, $(2^5 - 1)$ and $(2^7 - 1)$. Again someone saw a 'pattern,' claiming the next to be $(2^{11} - 1)$, but it is composite, being 23.89 (first noted in the 15th century).

Then I presented this problem (not the 'related' one of the title): Some natural numbers can/can't be expressed as a sum of at least two consecutive natural numbers; find some of each kind. Check down to 20. Soon they had two lists:

'can': 3, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20.

'can't': 1, 2, 4, 8, 16.

Now, who could guess the next entry in the ‘can’t’ list? There was a tentative suggestion that it might be 32. But is it correct? So we tested and found that 32 is in the ‘can’t’ list. And after 32? Is 64 the next one? And what about the ones in between, namely: 21, 22, 23, ..., 31, 33, ..., 63, 65, ...?

The problem now became: could they prove the conjectures:

- (i) any natural number that is NOT a power of 2 IS a sum of at least two consecutive natural numbers.
- (ii) any natural number that IS a power of 2 CANNOT be such a sum.

I didn’t expect that any of them would be able to prove these and indeed none could make a start. So I asked: how can one express notationally a sum of at least two consecutive natural numbers? Someone suggested $a + (a + 1)$, but it was quickly agreed that that wasn’t good enough and it was a while before we settled on: $a + (a + 1) + (a + 2) + \cdots + (a + b)$; $a, b \in N$.

Could they now prove conjecture (ii), that is:

is $2^m = a + (a + 1) + (a + 2) + \cdots + (a + b)$ impossible for $a, b, m \in N$ (and, trivially, $m = 0$)?

Could anyone tell me what $a + (a + 1) + \cdots + (a + b)$ summed to?, and after a while (which entailed going over high school work in connection with $1 + 2 + 3 + \cdots + b = b(b + 1)/2$, as some had said—from memory, and without reservation—that it came to $b(b + 1)$), we found the rearranged sum to be:

$$(a + a + a + \cdots + a) + (1 + 2 + \cdots + b) = a(b + 1) + b(b + 1)/2,$$

which comes to $(b + 1)(2a + b)/2$.

Could they now prove that:

$2^m = (b + 1)(2a + b)/2$, that is $2^{m+1} = (b + 1)(2a + b)$, is impossible for $a, b, m \in N$?

They were stuck again, and so I asked: if the product of two natural numbers is a power of 2, what can you tell me about those two natural numbers? Can anyone prove that the only factors of 2^n , $n \in N$, are 2^r , $r = 0, 1, 2, \dots, n$? And, while we’re at it, what can one say about the factors of $3^n, 4^n, 5^n, 6^n, \dots$?

It was seen that we really did have a problem here when they came up with examples like: 12 is a factor of 6, but 12 is not a power of 6; 2 is a factor of 4, but 2 is not a power of 4. Eventually—after much discussion, using only simple ideas, and not resorting to the unique factorization theorem (I will send details to anyone who is interested)—I proved that the only factors of 2^n are 2^r ($r = 0, 1, 2, \dots, n$), and a proof of conjecture (ii) easily followed:

Suppose $2^m = a + (a + 1) + \cdots + (a + b)$ for some $a, b, m \in N$, then $2^m = (b + 1)(2a + b)/2$, and so $2^{m+1} = (b + 1)(2a + b)$. Then $(b + 1)$ and $(2a + b)$ are 2^x and 2^y , some $x, y \in N$ (with, though it is not needed, $x + y = m + 1$)—observing that neither x nor y could be 0 as $(b + 1)$ and $(2a + b)$ are both greater than 1. Addition gives $(2a + 2b + 1) = 2^x + 2^y$, which is clearly impossible.

All of this had been quite a struggle, and so it was a relief that a solution to (i) came more quickly. First we had some discussion as to the shape of a natural

number that is NOT a power of 2—it must be like this: $2^c \cdot (2d + 1)$, c chosen from $0, 1, 2, \dots$; d from $1, 2, 3, \dots$. Now, (conjecture (i)) question:

Given c and d as above, are there a and $b \in N$, such that $2^c \cdot (2d + 1) = a + (a + 1) + (a + 2) + \dots + (a + b)$?, that is: $2^c \cdot (2d + 1) = (b + 1)(2a + b)/2$, which someone replaced by: $2^{c+1} \cdot (2d + 1) = (b + 1)(2a + b)$.

At first no one knew what to do, so I asked for values of c and d to be chosen; someone suggested $c = 4$ and $d = 7$, and I asked if anyone could find $a, b \in N$ with $2^5 \cdot 15 = (b + 1)(2a + b)$. Clare suggested choosing $b = 14$, making $b + 1 = 15$. Then? She also suggested setting $2a + b = 2^5$, which gives $a = 9$. Good! So, $2^4 \cdot 15 = (240) = 9 + 10 + 11 + \dots + 22 + 23$.

Another example? Someone choose $c = 4$ and $d = 2$, which led in identical manner to $2^4 \cdot 5 = (80) = 14 + 15 + 16 + 17 + 18$, but when I asked for another example with c held constant at 4 and d chosen differently, I was offered $d = 31$, and then Orla, with $2^5 \cdot 63$, suggested choosing $b + 1 = 63$, giving $b = 62$, and then $2^5 = 2a + b = 2a + 62$ gave $a = -15$. The uniform reaction of the class was to reject this value (“it’s not a natural number!”) and then Clare suggested not choosing $b + 1$ to be 63, but to be 2^5 . Good! That gave b to be 31, and then setting $2a + b = 63$ gave $a = 16$, and so:

$$2^4 \cdot 63 = (1008) = 16 + 17 + 18 + \dots + 46 + 47.$$

It then gave me great pleasure to point out that the ‘rejected’ value of $a(-15)$, together with the original $b(62)$, gave the same conclusion! $a = -15$, $b = 62$ give $a + b = 47$ and $2^4 \cdot 63 = -15 - 14 - \dots - 1 + 0 + 1 + \dots + 14 + 15 + 16 + \dots + 47$. Many of them seemed to be taken aback with that, but more examples made them feel at home with it and it was soon clear that either approach (with respect to the ‘choice’) now led to a proof of conjecture (i).

It’s easy to see that if ‘at least two’ is replaced by ‘three or more’ (Halmos’ problem) then only the primes from the above ‘can’ list transfer to the ‘can’t’ list, as $(b + 1)(2a + b)/2$ is clearly composite for $a, b \in N$, and $b > 1$ (‘three or more’); and every composite is so representable, and so stays in the ‘can’ list.

I then asked them to consider this question, the ‘related problem’ of the title: Some natural numbers can/can’t be expressed as a sum of at least two consecutive even/odd natural numbers. For example: $8 = 3 + 5$, $12 = 2 + 4 + 6$, etc.; but 7 and 13 are not so representable. Which can/can’t? Check to 20.

To the surprise of many these were the respective lists:

can: 4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20.

can’t: 1, 2, 3, 5, 7, 11, 13, 17, 19. (for them: 1 and the primes)

But could they prove the following conjectures (having discussed how one might represent a sum of consecutive even/odd natural numbers):

- (iii) if $m \in N$ and m is composite, are there $a, b \in N$ with $m = a + (a + 2) + (a + 4) + \dots + (a + 2b)$?
- (iv) if $p \in N$ and p is prime (or 1), is this impossible: $p = a + (a + 2) + (a + 4) + \dots + (a + 2b)$, all $a, b \in N$?

They were able to show that

$$\begin{aligned} a + (a + 2) + (a + 4) + \dots + (a + 2b) \\ &= (a + a + a + \dots + a) + (2 + 4 + \dots + 2b) \\ &= (b + 1)a + 2(1 + 2 + \dots + b) = (b + 1)a + 2 \cdot b(b + 1)/2 \\ &= (b + 1)a + b(b + 1) = (b + 1)(a + b), \end{aligned}$$

and then a proof of (iv) followed instantly. So did a proof of (iii), by the same approach as before—not smoothly, but at least not with as many holdups as earlier!

Finally, if you investigate which natural numbers can/can't be expressed as $a + (a + 3) + (a + 6) + \cdots + (a + 3b)$, $a, b \in N$, you will find (the latter list is complete down to 100):

can: 5, 7, 9, 11, 12, 13, 15, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 29, 30, ...

can't: 1, 2, 3, 4, 6, 8, 10, 14, 16, 20, 28, 32, 44, 52, 56, 64, 68, 76, 88, ...

Things are a bit more jumbled up than before, but here are some observations/exercises which I leave to the reader:

(0) (a trivial one) Every odd number from 5 is in the 'can' list (and so, of the primes, only 2 and 3 are in the other).

(1) Every power of 2 is in the 'can't' list (as with $c = 1$). In fact, if you consider which natural numbers can/can't be represented by general sums of the form:

$$a + (a + c) + (a + c2) + \cdots + (a + cb), \quad a, b, c \in N,$$

then NO power of 2 can be so represented when c is odd (the above cases are $c = 1$ and 3), but ALL powers of 2 can be so represented (apart from small, easily explained exceptions) when c is even.

(2) When $c = 1$ or 2 both 'can' lists are closed under product (the product of two non-powers of 2 is a non-power of 2, and the product of two composites is also composite). In fact the 'can' list is closed under product for all c . Prove it!

(3) There are many other structural questions that occur on examining the lists—and I don't wish to spoil your fun of finding your own—but just to mention one: consider g from a 'can't' list with $g > 1$.

$c = 1$: the product of two such g 's remains in the 'can't' list.

$c = 2$: the product of two such g 's goes to the 'can' list.

$c = 3$: here some products remain in the 'can't' list (e.g. $2 \cdot 3$, $2 \cdot 4$, $2 \cdot 8$, etc.) while others move to the 'can' list (e.g. $2 \cdot 6$, $3 \cdot 6$, $3 \cdot 3$, etc.). Prove (or disprove!) that if g is NOT a power of 2 then all proper powers of g are in the 'can' list. This is vacuous in the $c = 1$ case, is trivial in the $c = 2$ case, and is false in the $c = 4$ case (e.g. $3, 5, 7, 9, 25, 49 \in$ 'can't' list).

I wish to thank Professor Halmos for his very generous reply on receiving a first draft of this note, and record that in the problems section of his book he asks: which positive integers are sums of two or more consecutive integers?, and also that in the solutions section he asks (and answers): which positive integers are sums of two or more consecutive odd positive integers?, and: which positive integers are sums of three or more consecutive positive integers?

*Mathematics Department
St. Patrick's College
Dublin 9
Ireland*

NOTES

Edited by: John Duncan

A “Popular” Class Number Formula

Kurt Girstmair

1. A THEOREM ON THE DIGITS OF $1/p$. Even if one does not know what the class number is, he can understand the following *consequence* of our Theorem, which is quite deep (in the mathematical sense):

Let $p \geq 7$ be a prime number, $p \equiv 3 \pmod{4}$, and let

$$1/p = 0.x_1x_2x_3\dots$$

be the decimal representation of $1/p$. Thus $x_k \in \{0, 1, \dots, 9\}$ means the k th digit of $1/p$. This representation of $1/p$ is *recurring* (or *periodic*), i.e.,

$$1/p = 0.x_1\dots x_nx_1\dots x_nx_1\dots x_n\dots$$

holds for a certain natural number n . The smallest possible number n of this kind is called the *period length* of $1/p$. Suppose now that $1/p$ has the period length $n = p - 1$. This is the case, e.g., for $p = 7, 19, 23, 47, 59$, but not for $p = 11, 31, 43, 67$. Then

$$x_1 + x_3 + x_5 + \dots + x_{p-2} < x_2 + x_4 + \dots + x_{p-1}, \quad (1)$$

and the difference of these numbers is a multiple of 11.

Note that the sums in (1) are equal if $p \equiv 1 \pmod{4}$. This, however, is *not* deep and shown in Section 2. The inequality (1) means that an average digit of x_1, x_3, \dots, x_{p-2} is smaller than an average digit of x_2, x_4, \dots, x_{p-1} .

Example. Let $p = 47$. From

$$1/47 = 0.0212765957446808510638297872340425531914893617\dots$$

we get $x_1 + x_3 + \dots + x_{45} = 76$ and $x_2 + x_4 + \dots + x_{46} = 131$, whose difference is 55. Hence the average of x_1, x_3, \dots is $3\frac{7}{23}$, whereas the average of x_2, x_4, \dots is $5\frac{16}{23}$.

The Theorem holds for the digit expansion of $1/p$ with respect to an *arbitrary* basis $g \in \mathbb{Z}$, $g \geq 2$, provided that g is a primitive root mod p . The property of being primitive can be characterized as follows: For $k = 0, 1, 2, 3, \dots$ define $g_k \in \{0, 1, \dots, p - 1\}$ by

$$g_k \equiv g^k \pmod{p}.$$

Then g is primitive if, and only if, g_k runs through all of $1, \dots, p - 1$ if k does. The assumption “ g primitive” just means that the period length of the digit expansion of $1/p$ with respect to the basis g equals $p - 1$ (see [3]; [3] and [1] are standard references for digit expansions).

The most important notion in the Theorem is the *class number* h of the discriminant $-p$ (which is identical with the class number of the field $\mathbb{Q}(\sqrt{-p})$). A definition of h is given in Section 2. For the reader not interested in details it

may suffice that h is a positive integer and one of the fundamental number theoretic data attached to p . That is why number theorists need a formula to compute h for a given p . The Theorem supplies such a “class number formula.”

Theorem. Let $p \geq 7$ be a prime, $p \equiv 3 \pmod{4}$, and g a primitive root mod p . Let

$$\frac{1}{p} = \sum_{k=1}^{\infty} x_k g^{-k}, \quad x_k \in \{0, 1, \dots, g-1\}, \quad (2)$$

be the (uniquely determined) digit expansion of $1/p$ with respect to the basis g . Let h be the class number of the discriminant $-p$. Then

$$\sum_{k=1}^{p-1} (-1)^k x_k = (g+1)h.$$

In the case $g = 10$ the Theorem says

$$(x_2 + x_4 + \dots + x_{p-1}) - (x_1 + x_3 + \dots + x_{p-2}) = 11 \cdot h,$$

whence (1) follows. We shall see that (2) is a transformation of a famous class number formula of G. P. L. Dirichlet (1805–1859). In order to enounce Dirichlet’s formula, we need the *Legendre symbol*

$$\left(\frac{k}{p}\right)$$

of an integer k not divisible by p . By its definition, $\left(\frac{k}{p}\right) = 1$, if k is a square mod p , i.e., $k \equiv j^2 \pmod{p}$ for some $j \in \mathbb{Z}$; otherwise $\left(\frac{k}{p}\right) = -1$. The Legendre symbol has the following properties, which will be used in the proof of the Theorem (they are easy to prove, see [4], p. 35 f.):

- (A) For a primitive root $g \pmod{p}$, $\left(\frac{g}{p}\right) = -1$.
- (B) If $k \equiv j \pmod{p}$, $\left(\frac{k}{p}\right) = \left(\frac{j}{p}\right)$.
- (C) For all k and j (not divisible by p), $\left(\frac{kj}{p}\right) = \left(\frac{k}{p}\right)\left(\frac{j}{p}\right)$.

Now Dirichlet’s class number formula says: For each prime $p \equiv 3 \pmod{4}$, $p \geq 7$,

$$h = - \sum_{k=1}^{p-1} \left(\frac{k}{p}\right) \frac{k}{p}. \quad (3)$$

In Section 2 we give some indication of the ingredients of the proof of (3). A complete proof can be found, e.g., in [4], p. 300 ff. Since it is a class number, the expression on the right side of (3) must be positive. There seems to be no other way to see this (except in special cases). That is why the positivity of the right side of (3) belongs to the deep results of number theory; and so does (1), which is equivalent to this result in an elementary manner.

Proof of the Theorem: Since g is primitive,

$$\sum_{k=1}^{p-1} \left(\frac{k}{p}\right) \frac{k}{p} = \sum_{k=1}^{p-1} \left(\frac{g_k}{p}\right) \frac{g_k}{p}.$$

By the properties (A)–(C) of the Legendre symbol we obtain

$$\left(\frac{g_k}{p}\right) = \left(\frac{g^k}{p}\right) = \left(\frac{g}{p}\right)^k = (-1)^k.$$

For these reasons (3) can be written as

$$h = \sum_{k=1}^{p-1} (-1)^{k+1} \frac{g_k}{p}.$$

Now

$$(g+1)h = \sum_{k=1}^{p-1} (-1)^{k+1} \frac{gg_k}{p} - \sum_{k=1}^{p-1} (-1)^k \frac{g_k}{p} = \sum_{k=1}^{p-1} (-1)^k \frac{gg_{k-1} - g_k}{p}.$$

For integers $k \geq 1$ put

$$y_k = \frac{gg_{k-1} - g_k}{p}.$$

The proof is complete if we can show that $y_k = x_k$ for all k . Indeed, y_k is an integer since $gg_{k-1} \equiv g^k \equiv g_k \pmod{p}$. It is in the range $0, \dots, g-1$. For, on the one hand, $g_{k-1}/p < 1$ yields $y_k < g$; on the other hand, $-g_k/p > -1$ yields $y_k > -1$. Finally,

$$\sum_{k=1}^{\infty} y_k g^{-k} = \frac{1}{p} \left(\sum_{k=1}^{\infty} \frac{g_{k-1}}{g^{k-1}} - \sum_{k=1}^{\infty} \frac{g_k}{g^k} \right) = \frac{1}{p}.$$

The uniqueness of the digit expansion shows $y_k = x_k$.

For generalizations of the Theorem and other results about the digits of $1/p$ see [2].

2. REMARKS ON THE THEOREM. Clearly the identity

$$x_k = \frac{gg_{k-1} - g_k}{p} \tag{4}$$

also holds for a prime $p \equiv 1 \pmod{4}$ if g is a primitive root mod p . We show that in this case

$$x_1 + x_3 + \dots + x_{p-2} = x_2 + x_4 + \dots + x_{p-1} = (g-1)(p-1)/2.$$

Let $n = (p-1)/2$. Since $g^{p-1} \equiv 1 \pmod{p}$, $g^n \equiv \pm 1 \pmod{p}$. But $g^n \equiv 1 \pmod{p}$ contradicts the primitivity of g , so $g^n \equiv -1 \pmod{p}$. Therefore $g_{n+k} \equiv -g_k \pmod{p}$, which implies $g_{n+k} = p - g_k$ for all $k \geq 0$. From (4) we obtain

$$x_{n+k} = \frac{g(p - g_{k-1}) - (p - g_k)}{p} = g - 1 - x_k, \quad k \geq 1,$$

i.e., $x_k + x_{n+k} = g - 1$. Because of $p \equiv 1 \pmod{4}$ the number n is even, and

$$\begin{aligned} x_1 + x_3 + \dots + x_{p-2} &= (x_1 + x_{n+1}) + (x_3 + x_{n+3}) \\ &\quad + \dots + (x_{n-1} + x_{2n-1}) = (g-1) \cdot \frac{n}{2}. \end{aligned}$$

The same argument works for $x_2 + x_4 + \dots + x_{p-1}$.

Next we define the class number h of the discriminant $-p$, $p \equiv 3 \pmod{4}$. We consider *integral binary quadratic forms* (henceforth simply called *forms*). By this we mean polynomials of the shape

$$f = aX^2 + bXY + cY^2$$

with coefficients $a, b, c \in \mathbb{Z}$. Two forms f, f' are said to be *equivalent*, if there is a matrix

$$A = \begin{pmatrix} r & s \\ t & u \end{pmatrix}$$

with integral entries and determinant $ru - st = 1$ such that

$$f' = f(rX + sY, tX + uY).$$

Since the inverse of A is of the same type (integral entries, determinant = 1), the equivalence of forms is well defined; it is an equivalence relation, indeed. The number $D(f) = b^2 - 4ac$ is called the *discriminant* of f . The discriminant remains invariant under equivalence: If f and f' are equivalent, $D(f) = D(f')$. Therefore it is quite natural to consider the equivalence classes of forms with a *fixed* discriminant. In the sequel we always assume that $D(f) = -p$, p a prime, $p \equiv 3 \pmod{4}$. There is at least one form with this property, namely

$$f_0 = X^2 + XY + \frac{p+1}{4}Y^2.$$

Moreover, if $f = aX^2 + bXY + cY^2$ and $f' = a'X^2 + b'XY + c'Y^2$ are equivalent, their first coefficients a and a' have the same sign. Thus the following definition of the class number h makes sense:

h is the number of equivalence classes of forms with discriminant $-p$ and first coefficient > 0 .

Since f_0 is such a form, h cannot be 0, so $h \geq 1$. It is not trivial, however, that h is finite.

Dirichlet was able to connect h with the convergent series

$$\sum_{k=1}^{\infty} \left(\frac{k}{p} \right) \frac{1}{k}.$$

A thorough analysis of quadratic forms and a geometric argument enabled him to show for $p \geq 7$

$$\sum_{k=1}^{\infty} \left(\frac{k}{p} \right) \frac{1}{k} = \frac{\pi h}{\sqrt{p}}. \quad (5)$$

On the other hand he proved

$$\sum_{k=1}^{\infty} \left(\frac{k}{p} \right) \frac{1}{k} = -\frac{\pi}{\sqrt{p}} \sum_{k=1}^{p-1} \left(\frac{k}{p} \right) \frac{k}{p}. \quad (6)$$

From (5) and (6) formula (3) follows immediately. The proof of (6) involves complex analysis (Abel's limit theorem) and the identity

$$\sum_{k=1}^{p-1} \left(\frac{k}{p} \right) e^{2\pi i k / p} = +i\sqrt{p}.$$

The difficult part in the proof of this identity is the sign of the right side. It took Gauss four years to show that the sign is always “+.”

REFERENCES

1. L. E. Dickson, *History of the Theory of Numbers*, vol. I (reprint), Chelsea Publ. Comp., New York 1952.
2. K. Girstmair, The digits of $1/p$ in connection with class number factors, to appear in *Acta Arith.*
3. G. H. Hardy, E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press, Oxford 1954.
4. L. K. Hua, *Introduction to Number Theory*, Springer-Verlag, Berlin 1982.

*Institut für Mathematik, Universität Innsbruck
Technikerstr. 25 / 7, A-6020 Innsbruck, Austria
Kurt.Girstmair@uibk.ac.at*

Variations on Wolstenholme's Theorem

Emre Alkan

Wolstenholme's Theorem is stated and proved in [1]. Our aim is to give similar results.

Theorem 1 (Wolstenholme). *If p is a prime greater than 3, then the numerator of the fraction $1 + \frac{1}{2} + \frac{1}{3} + \cdots + 1/(p-1)$ is divisible by p^2 .*

We first give an equivalent theorem in the sense that Wolstenholme's Theorem can be deduced from it.

Theorem 2. *If $p > 3$, then the numerator of*

$$\frac{1}{1(p-1)} + \frac{1}{2(p-2)} + \cdots + \frac{1}{\left(\frac{p-1}{2}\right)\left(\frac{p+1}{2}\right)}$$

is divisible by p .

Theorem 3. *If $p > 3$, then the numerator of $1 + 1/2^2 + \cdots + 1/(p-1)^2$ is divisible by p .*

(Mod p); so these congruences reduce to the set of congruences below.

$$\begin{array}{lll} 1. x & \equiv (-1)^{\phi(m)/(p-1)} (\text{Mod } p) \\ 2. x & \equiv (-1)^{\phi(m)/(p-1)} (\text{Mod } p) \\ \vdots & & \vdots \\ (p-1)x & \equiv (-1)^{\phi(m)/(p-1)} (\text{Mod } p) \end{array}$$

The solutions are clearly distinct (Mod p). Hence, they are $1, 2, \dots, p-1$ in some order. Finally, we have:

$$C_p \equiv \frac{\phi(m)}{p-1} (1 + 2 + \dots + p-1) \equiv \frac{\phi(m)}{p-1} \cdot \frac{p-1}{2} \cdot p \equiv 0 \pmod{p},$$

where C_p is the numerator of the fraction.

ACKNOWLEDGMENT. I wish to thank C. K. Bayram for his help and encouragement in the preparation of this note.

REFERENCE

1. G. H. Hardy, E. M. Wright, *An Introduction to the Theory of Numbers*, Fifth Edition, 1979.

(Undergraduate Student)
Department of Mathematics
Bosphorus University
Istanbul,
Turkey

The Second-Partials Test for Local Extrema of $f(x, y)$

Leonard Gillman

The second-partials test for local extrema provides a sufficient condition that a critical point of $f(x, y)$ be a local minimum (or a local maximum). Every college calculus book presents this test. This note concerns the proof. The standard proof is straightforward, and is the one usually given. Some texts however present an “advanced calculus” argument, containing epsilonic manipulations beyond the comfort level of most freshmen. Finally, there are texts that omit a proof altogether, as being a topic for advanced calculus. In a “random” sample of a dozen freshman texts (i.e., those within ten feet of my desk), the three types of presentation were about equally divided. The two types of proof also appeared equally in my half-dozen advanced calculus texts.

I believe that the standard proof is appropriate to the course and moreover furnishes an enlightening view of mathematics at work.

1. THE ONE-VARIABLE CASE. For background, start with one variable. Assume that $f(x)$ is twice differentiable on an open interval about a critical point $x = a$.

There are two versions of the test, which *faute de mieux* I'll call geometric and analytic.

Geometric. The standard second-derivative test argues that if $f''(x) > 0$ on the interval then f' is increasing there, so f is concave up and a is a local minimum.

Analytic. This version, not always mentioned, is based on Taylor's formula for $n = 1$. Since f' vanishes at a , the formula reduces to

$$f(a + h) - f(a) = \frac{1}{2}h^2f''(a + \theta h), \quad 0 < \theta < 1.$$

It is then obvious that a is a local minimum.

2. THE TWO-VARIABLE CASE. The test for a local minimum of $f(x, y)$ at a critical point (a, b) depends on the following fact:

$$\begin{aligned} AX^2 + 2BXY + CY^2 \text{ is positive definite provided that} \\ \Delta > 0 \text{ and } A > 0, \text{ where} \end{aligned} \quad (1)$$

$$\Delta = AC - B^2.$$

(To prove (1), complete the square to write the form as $[(AX + BY)^2 + \Delta Y^2]/A$.)

In the test itself, A , B , and C are given by

$$A = f_{11}(x, y), \quad B = f_{12}(x, y), \quad C = f_{22}(x, y),$$

assumed continuous on a neighborhood of (a, b) . The condition for a local minimum is that the inequalities

$$A > 0 \quad \text{and} \quad \Delta > 0$$

hold at (a, b) . By continuity, they then hold on some open disk D about (a, b) . Again there are two versions of the test. This time the analytic version is the popular one.

Analytic. Since $f_1(a, b) = f_2(a, b) = 0$, Taylor's formula for $n = 1$ reduces to

$$f(a + h, b + k) - f(a, b) = \frac{1}{2}[Ah^2 + 2Bhk + Ck^2], \quad (2)$$

where A , B , and C are evaluated at some point $(a + th, b + tk)$, $0 < t < 1$. Any such point lies in the disk D whenever $(a + h, b + k)$ does. By (1), the right side of (2) is positive for all points $(a + h, b + k) \neq (a, b)$ in D , and therefore f has a local minimum at (a, b) .

How can such a nice proof get to be complicated? The standard way seems to be to use the form $\varphi(x, y) + R$ instead of $\varphi(a + \theta h, b + \theta k)$ and then get involved making R small. The epsilonics are not only forbidding in their own right but add twenty lines to the argument. No student is going to struggle through that.

Geometric. This version casts the problem in terms of a function of a single variable. Introduce polar coordinates $\langle r, \alpha \rangle$ relative to (a, b) :

$$(x, y) = (a + r \cos \alpha, b + r \sin \alpha),$$

and define

$$F_\alpha(r) = f(x, y) = f(r \cos \alpha, r \sin \alpha).$$

Now differentiate twice with respect to r :

$$F'_\alpha(r) = f_1(x, y)\cos \alpha + f_2(x, y)\sin \alpha,$$

$$F''_\alpha(r) = A \cos^2 \alpha + 2B \cos \alpha \sin \alpha + C \sin^2 \alpha.$$

(These are $\nabla_\alpha f$ and $\nabla_\alpha \nabla_\alpha f$, the first and second directional derivatives in the direction α .)

Since the sine and cosine are never zero together, (1) shows that $F''_\alpha(r) > 0$ at all points of D . Then the trace of f in every direction from (a, b) is concave up throughout D . Therefore f has a local minimum at (a, b) .

The derivation of the two-variable Taylor expansion also starts out with a single variable, setting $F(t) = f(a + th, b + tk)$ and differentiating twice; and in fact, except for notation, this is identical to finding $F''_\alpha(r)$. Thus in the two-variable setting, the analytic and geometric proofs are very closely related. Still, I think the geometric version has a certain charm to it and should be better known.

ACKNOWLEDGMENT. I am indebted to the *Notes* editor for a helpful mix of prodding and advice.

1606 The High Road
Austin TX 78746
len@math.utexas.edu

PICTURE PUZZLE

(from the collection of Paul Halmos)



This is not an algebraic Hawaiian picture but an analytic one.
(see page 1012.)

NITSA MOYSHOVITZ-HADAR got her B.Sc. from the Hebrew University in Jerusalem, her M.Sc. from Technion–Israel Institute of Technology, and her Ph.D. from the University of California at Berkeley in 1975. Since then she has been on the faculty at Technion as a mathematics educator. She has recently become the academic director of The National Pedagogical Center for Mathematics which she initiated and helped establish. Her main interest is in the role of intuition and of counter-intuitive phenomena in the learning of mathematics.

RICHARD A. BRUALDI received a Ph.D. from Syracuse University in 1964 and has been on the faculty of the University of Wisconsin–Madison since 1965. In 1986 he was the recipient of the ‘Chancellor’s Award for Excellence in Teaching.’ His research interests include combinatorics, matrix theory and coding theory. He is the author of the book ‘Introductory Combinatorics’ (1st edition 1977, 2nd edition 1991) and coauthor (with H. J. Ryser) of the book ‘Combinatorial Matrix Theory’ (1991). He is a former editor of the Notes section of the *Monthly* (1975–78) and is currently co-editor-in-chief of the journal *Linear Algebra and its Applications*. He has been a vegetarian for almost 25 years, and even during the cold Wisconsin winters he can be spotted jogging along the shores of frozen Lake Monona in Madison.

STEPHEN MELLENDORF received a B.S. from Michigan State University in 1989. He is currently a graduate student at the University of Wisconsin–Madison working on his Ph.D. dissertation under the direction of Richard A. Brualdi. His research interests include graph theory, matrix theory and coding theory. He enjoys playing many sports including basketball and table tennis.

BLAIR K. SPEARMAN completed his B.Sc. and M.Sc. degrees at Carleton University in Ottawa, Ontario, Canada. He received his Ph.D. in mathematics at Pennsylvania State University under W.C. Waterhouse in 1981. He currently teaches at Okanagan University College, Kelowna, B.C., Canada. His research interests are in algebraic number theory.

KENNETH S. WILLIAMS received his Ph.D. degree in mathematics from the University of Toronto in 1965 under J.H.H. Chalk. After spending the year 1965–66 at the University of Manchester, England, he joined the faculty of Carleton University, where he is currently Professor of Mathematics. In 1979 he was awarded a D.Sc. degree by the University of Birmingham, England. He served as the Chairman of the Department of Mathematics and Statistics at Carleton University from 1980 to 1984. His research interests are in number theory (algebraic, analytical and computational). In his spare time (when he has any) he enjoys running, gardening and walking is two golden retrievers Newton and Zena.

JOHN B. COSGRAVE studied at Royal Holloway College of London University (B.Sc.(1968) and Ph.D.(1972)), taught in its Mathematics Department for four years and has fond memories of its Head, the great convexity expert, H. G. Eggleston. Later he taught in Manchester Univ. (U.K.), Ibadan Univ. (Jos campus, Nigeria), Carysfort College (Dublin), and St. Patrick’s College (Dublin). His first mathematical love is elementary number theory.

STEPHEN B. MAURER (B.A. Swarthmore 1967, Ph.D. Princeton 1972) has written and spoken widely on discrete mathematics and curriculum. His research has been in combinatorics, with forays (sometimes continuous) into mathematical biology, economics and anthropology. His article “The King Chicken Theorems” won the MAA’s 1981 Allendoerfer Award for expository writing. He has been a program officer at the Sloan Foundation and chaired the MAA’s high school competitions. He is a “Math Dad” volunteer for his older kid’s second grade class.

Answer to Picture Puzzle (p. 1006)

Richard Courant, Peter Lax, and Michael Atiyah relaxing in Honolulu
in 1969.

UNSOLVED PROBLEMS

Edited by: **Richard Guy and Richard Nowakowski**

In this department the MONTHLY presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial or related results. Typescripts should be sent to Richard Guy, Department of Mathematics & Statistics, The University of Calgary, Alberta, Canada T2N 1N4.

Mousetrap

Richard K. Guy and Richard J. Nowakowski

Cayley [2, 3] introduced a permutation problem he called *Mousetrap* which is loosely based on the card game Treize [1]. Suppose that the numbers $1, 2, \dots, n$ are written on cards, one to a card. After shuffling (permuting) the cards, start counting the deck from the top card down. If the number on the card does not equal the count, transfer the card to the bottom of the deck and continue counting. If the two are equal then set the card aside and start counting again from 1. The game is *won* if all the cards have been set aside, but lost if the count reaches $n + 1$.

For example the permutation 1243 will win, the cards being set aside in the order 1, 3, 4, 2, while 1324 only sets aside cards 1 and 2.

Cayley proposed two questions.

1. For each n find all the winning permutations of $1, 2, \dots, n$.
2. For each n find the number of permutations that eliminate precisely i cards for each i , $1 \leq i \leq n$.

An answer for question 2 would illuminate question 1, but very little is known about either.

In [3], Cayley lists all the possible outcomes for $n = 4$. Steen [7] notes that Cayley made some errors. He goes on to calculate, for any n , the number of permutations that have i , $1 \leq i \leq n$, as the first card set aside. This number he denotes by $a_{n,i}$ and he used $b_{n,i}(c_{n,i})$ to denote the number of permutations that have 1 (respectively 2) as the first hit and i as the second. He obtained the recurrence relations

$$a_{n,i} = a_{n,i-1} - a_{n-1,i-1}, \quad b_{n,i} = a_{n-1,i-1}$$
$$c_{n,i} = c_{n,1} - (i-1)c_{n-1,1} + \sum_{k=2}^{i-2} (-1)^k \frac{i(i-1-k)}{2} c_{n-k,1} \quad \text{for } n > i+1$$

and used them to show that for $0 \leq i \leq n$,

$$a_{n,0} = na_{n-1,0} + (-1)^n, \quad a_{0,0} = 1; \quad a_{n,i} = \sum_{k=0}^i (-1)^k \binom{i}{k} (n-1-k)!$$

$$b_{n,i} = a_{n-1,i-1} = a_{n-2,i-2} - a_{n-3,i-2}$$

$$c_{n,i} = \sum_{k=1}^{i-3} (-1)^{k+i-1} \frac{k(k+3)}{2} (n-i+k-1)! - (i-1)(n-3)! + (n-2)!$$

Steen denoted the sums of the $a_{n,i} b_{n,i} c_{n,i}$ taken over $0 \leq i \leq n$ (but omitting $i = 0, i = 1, i = 2$ respectively) by $a_n b_n c_n$ and further showed that

$$a_n = na_{n-1} + (-1)^{n+1} \quad b_n = a_{n-1}$$

and deduced a complicated expression for c_n from his formula for $c_{n,i}$ which unfortunately holds for neither $i = n$ nor $i = n - 1$. In [4] there are corrected versions of Steen's formulas:

$$c_{n,n-1} = \sum_{k=0}^{n-3} (-1)^k \binom{n-3}{k} (n-k-2)!$$

$$c_{n,n} = (n-2)! + \sum_{k=0}^{n-5} (-1)^{k+1} \left(\binom{n-4}{k} + \binom{n-3}{k+1} \right) (n-k-3)! + 2(-1)^{n-3}$$

$$c_n = (n-2)(n-2)! - \left[\frac{1}{e} ((n-1)! - (n-2)! - 2(n-3)!) \right],$$

where $[[x]]$ is the nearest integer to x .

If $i = 0$, so that no card is set aside then the permutation is a *derangement* ([6], sequence 766), that is, card j is in place j for no value of j . Sloane [6] also contains some related sequences. Steen's a_n, c_n and $(n-2)(n-2)! - c_n$ are sequences 1423, 1186 and 1635, respectively. The first edition of [6] cited Steen's sometimes erroneous values of the last two for $3 \leq n \leq 10$. They have been corrected in the forthcoming second edition. They can be calculated from the formula above.

Notice that the number of permutations which eliminate at least one card is the nearest integer to $n!(1 - \frac{1}{e})$. This is the special case $j = n - 1$ of a more general problem [5]: find the minimum number of permutations of $1, 2, \dots, n$ which contain all permutations of a given j -element subset. More generally for $n \geq 2k - 1$ there is an expression of the form

$$\left[\frac{1}{e} ((n-1)! - (k-1)(n-2)! + (k^2 - 5k + 2)(n-3)! + \dots + (-1)^{k-1} (2k-3)(n-2k+2)! + (-1)^{k-1} 2(n-2k+1)!) \right].$$

A third question arose during our investigations about which we also know very little. Consider a permutation for which every number is set aside. The list of numbers in the order that they were set aside is another permutation. Any permutation obtained in this way we call a *reformed* permutation.

3. Characterize the reformed permutations.

Not all permutations are reformed permutations. For example, permutations on n objects are not reformed permutations if they start with n ; or with $x, n-1, y$ where $y \neq n \neq x$. On the other hand, the identity permutation is always a reformed permutation; the permutations yielding this are 1, 12, 132, 1423, 13254, 142563, 1527436, 16245378, 142863795, ...

The permutation 4213 is a winning permutation which gives rise to the permutation 2134; this in turn gives the reformed permutation 3214 which is not a winning permutation.

4. For a given n , what is the longest sequence of reformed permutations?

Table 1, whose column sums are $n!$, gives the numbers of permutations yielding sequences of length l .

TABLE 1. NUMBERS OF SEQUENCES OF REFORMED PERMUTATIONS.

l	$n =$	1	2	3	4	5	6	7	8	9
0		1	1	4	18	105	636	4710	38508	352902
1		—	1	2	4	14	72	316	1730	9728
2					2	1	11	14	81	242
3							1		1	8

5. Are there sequences of arbitrary length? Are there any cycles other than

$$1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \dots \quad \text{and} \quad 12 \rightarrow 12 \rightarrow 12 \rightarrow 12 \dots ?$$

MODULAR MOUSETRAP. We can play Mousetrap, but instead of counting $n, n+1, \dots$, we can start again, $\dots, n, 1, 2, \dots$. Now at least as many cards get set aside. In fact if n is prime, then either the initial deck is a derangement, or all cards get set aside, so every sequence cycles or terminates in a derangement. The identity permutation $123 \dots n$ will always form a 1-cycle and now there are also examples of nontrivial cycles.

For $n = 2$, $12 \rightarrow 12 \rightarrow 12 \rightarrow \dots$ cycles and 21 terminates.

For $n = 3$, $132 \rightarrow 123 \rightarrow 123 \rightarrow \dots$ cycle, while $321 \rightarrow 213 \rightarrow 312$ and 231 terminate.

If n is composite, the number of cards set aside may be strictly between 0 and n . As before, exactly $n-1$ cards cannot be set aside; and it's easy to see that neither can just one card. For example, with $n = 4$ there are 9 derangements with the permutations 2431 and 4132 at distance 1 from two of them; 7 permutations which set aside just 2 cards and 4213 at distance 1 from one of them; and the 1-cycle 1234 with two permutations at distance 1 from it and 1243 and 1432 at distance 2.

For $n = 5$ there is a 2-cycle, 21345 and 32145; and there is also a path of length 4:

$$54321 \rightarrow 34215 \rightarrow 52143 \rightarrow 21435 \rightarrow 51423.$$

6. Are there k -cycles for every k ? What is the least value of n which yields a k -cycle?

We are grateful to Sherwood Washburn for bringing Mousetrap to our attention and for supplying copies of the early literature.

REFERENCES

1. W. W. Rouse Ball and H. S. M. Coxeter, *Mathematical Recreations and Essays* 12th edition, Univ. of Toronto, 1974, pp. 336–337.
2. A. Cayley, A Problem in Permutations, *Quart. Math. J.*, I (1857), 79.
3. A. Cayley, On the Game of Mousetrap, *Quart. J. Pure Appl. Math.*, XV (1877), 8–10.
4. Richard K. Guy and Richard J. Nowakowski, Mousetrap, Proc. Erdős 80 Kesthely Combin. Conf., 1993.
5. Peter J. Slater, How few n -permutations contain all possible k -permutations? *Amer. Math. Monthly*, 90 (1983) 461.
6. N. J. A. Sloane, *A Handbook of Integer Sequences*, Academic Press, 1973.
7. Adolf Steen, Some Formulae Respecting the Game of Mousetrap, *Quart. J. Pure Appl. Math.*, XV (1878), 230–241.

Department of Mathematics
University of Calgary
Calgary, Alberta
Canada T2N 1N4

Department of Mathematics
Dalhousie University
Halifax, Nova Scotia
Canada B3H 4H8

December and June

cold
winds howl
geese go south
nights long June waits
temperatures fall low
ponds freeze snowmen grow
toboggans slide down hillsides
sun hides ice coats June waits
wood-fires flame snowballs fly
winds howl groundhogs hibernate

sun glows raspberries ripen
catbird sings iris blooms
days bright streams play June dreams
holiday picnics catch flies
wheat thrives crickets chirp
tomato plants climb
streams dance June plays
catbird sings
sun glows
warm

From *Intersections: Poems by JoAnne Growney*,
Kadet Press, Bloomsburg, PA, 1993, p. 50.

*The numbers of syllables in the phrases of this poem
follow the patterns of factorization of the integers from 1
to 10, then 10 to 1, into prime factors.*

THE AUTHORS

EDWARD SCHEINERMAN received his bachelor's from Brown University and doctorate from Princeton University. He is Professor in the Department of Mathematical Sciences—and has a joint appointment in Computer Science—at Johns Hopkins. He serves as a managing editor for the *Journal of Graph Theory*.

ORA ENGELBERG PERCUS received her Ph.D. in Mathematical Statistics from Columbia University in 1965. She has been active in applied probability, combinatorics, discrete queueing networks, and numerical simulations at the Courant Institute of Mathematical Sciences, New York University.

JEROME K. PERCUS received his Ph.D. in Theoretical Physics from Columbia University in 1954. He teaches at the Courant Institute and the Physics Department at New York University, and dabbles in mathematical physics and mathematical biology in his spare time.

At an early age, **DON PAUL RAWLINGS** gave up cowboying and the chance of a career on his parents' Arizona ranch to pursue mathematics. He wrote his doctoral thesis at the University of Strasbourg under the direction of Dominique Foata and received a Ph.D. in mathematics from the University of California at San Diego in 1979. He then took up residence at California Polytechnic State University in San Luis Obispo. Besides a certain obsession for combinatorics and permutation statistics, Don enjoys picking the guitar and clogging (a form of tap dance).

WILLIAM PAULSEN received his Ph.D. from Washington University in 1990 for work on the Hausdorff dimension of fractals. He has since changed his main research interest to applied mathematics, and is currently teaching at Arkansas State University in Jonesboro. His current project is computing the eigenfrequencies of various beam structures.

PETER HILTON is Distinguished Professor of Mathematics Emeritus, at SUNY, Binghamton. He is the author of numerous books and research articles on algebraic topology, homological algebra, and group theory. He has a long-standing interest in mathematics education—he was Chairman of the United States Commission on Mathematical Instruction and of the National Research Council Committee on Applied Mathematics Training. He was recently invited speaker at the “Georges de Rham Day” at the University of Lausanne, Switzerland. In May, 1993, an International Conference on Algebra and Topology was held at the Centre de Recherche Mathématique, Université de Montréal, to mark his 70th birthday.

JEAN PEDERSEN is Associate Professor of Mathematics at Santa Clara University. She is a past Governor of the Northern California Section of the MAA. In 1992 she received an award from her own university for “Outstanding achievement in teaching, research, and service to the Department of Mathematics.” She is the author of several articles and books on geometry—and has been much concerned, with Peter Hilton, to revive the study of geometry at all levels and to develop an integrated mathematics curriculum. They are joint authors of *Fear No More: An adult approach to mathematics, Build your own polyhedra*—and, with Jean Benson, *College Preparatory Mathematics*. They are currently engaged, with Derek Holton, in preparing *Mathematics: Reflections in a room with many mirrors*, a collection of tempting mathematical tidbits.

ISRAEL KLEINER is professor of mathematics at York University in Toronto. He received his Ph.D. in ring theory from McGill University. His current research interests are the history of mathematics, mathematics education, and especially their interface. Professor Kleiner is the coordinator of an in-service Master's Programme for teachers of mathematics, and has been involved in liaison work with the schools for many years.

NITSA MOYSHOVITZ-HADAR got her B.Sc. from the Hebrew University in Jerusalem, her M.Sc. from Technion–Israel Institute of Technology, and her Ph.D. from the University of California at Berkeley in 1975. Since then she has been on the faculty at Technion as a mathematics educator. She has recently become the academic director of The National Pedagogical Center for Mathematics which she initiated and helped establish. Her main interest is in the role of intuition and of counter-intuitive phenomena in the learning of mathematics.

RICHARD A. BRUALDI received a Ph.D. from Syracuse University in 1964 and has been on the faculty of the University of Wisconsin–Madison since 1965. In 1986 he was the recipient of the ‘Chancellor’s Award for Excellence in Teaching.’ His research interests include combinatorics, matrix theory and coding theory. He is the author of the book ‘Introductory Combinatorics’ (1st edition 1977, 2nd edition 1991) and coauthor (with H. J. Ryser) of the book ‘Combinatorial Matrix Theory’ (1991). He is a former editor of the Notes section of the *Monthly* (1975–78) and is currently co-editor-in-chief of the journal *Linear Algebra and its Applications*. He has been a vegetarian for almost 25 years, and even during the cold Wisconsin winters he can be spotted jogging along the shores of frozen Lake Monona in Madison.

STEPHEN MELLENDORF received a B.S. from Michigan State University in 1989. He is currently a graduate student at the University of Wisconsin–Madison working on his Ph.D. dissertation under the direction of Richard A. Brualdi. His research interests include graph theory, matrix theory and coding theory. He enjoys playing many sports including basketball and table tennis.

BLAIR K. SPEARMAN completed his B.Sc. and M.Sc. degrees at Carleton University in Ottawa, Ontario, Canada. He received his Ph.D. in mathematics at Pennsylvania State University under W.C. Waterhouse in 1981. He currently teaches at Okanagan University College, Kelowna, B.C., Canada. His research interests are in algebraic number theory.

KENNETH S. WILLIAMS received his Ph.D. degree in mathematics from the University of Toronto in 1965 under J.H.H. Chalk. After spending the year 1965–66 at the University of Manchester, England, he joined the faculty of Carleton University, where he is currently Professor of Mathematics. In 1979 he was awarded a D.Sc. degree by the University of Birmingham, England. He served as the Chairman of the Department of Mathematics and Statistics at Carleton University from 1980 to 1984. His research interests are in number theory (algebraic, analytical and computational). In his spare time (when he has any) he enjoys running, gardening and walking is two golden retrievers Newton and Zena.

JOHN B. COSGRAVE studied at Royal Holloway College of London University (B.Sc.(1968) and Ph.D.(1972)), taught in its Mathematics Department for four years and has fond memories of its Head, the great convexity expert, H. G. Eggleston. Later he taught in Manchester Univ. (U.K.), Ibadan Univ. (Jos campus, Nigeria), Carysfort College (Dublin), and St. Patrick’s College (Dublin). His first mathematical love is elementary number theory.

STEPHEN B. MAURER (B.A. Swarthmore 1967, Ph.D. Princeton 1972) has written and spoken widely on discrete mathematics and curriculum. His research has been in combinatorics, with forays (sometimes continuous) into mathematical biology, economics and anthropology. His article “The King Chicken Theorems” won the MAA’s 1981 Allendoerfer Award for expository writing. He has been a program officer at the Sloan Foundation and chaired the MAA’s high school competitions. He is a “Math Dad” volunteer for his older kid’s second grade class.

Answer to Picture Puzzle (p. 1006)

Richard Courant, Peter Lax, and Michael Atiyah relaxing in Honolulu
in 1969.

PROBLEMS AND SOLUTIONS

Edited by:

Richard T. Bumby, Fred Kochman and Douglas B. West

Proposed problems should be sent to the MONTHLY PROBLEMS address given on the inside front cover. Please include solutions and relevant references. Three copies of all items needed to evaluate the problem should be sent.

Solutions of published problems should arrive before May 31, 1995 at the MONTHLY PROBLEMS address given on the inside front cover. If possible, solutions should be typed with double spacing. Two copies suffice. Several solutions may be mailed together, but they should be on separate sheets of paper. The problem number and the solver's name and mailing address should appear on each solution. A mailing label should be included if an acknowledgment is desired.

The published solution is likely to be based on a solution that is complete and correct. Additional information, such as references to other appearances of the problem or its solution, is also welcome.

An asterisk () after the number of a problem, or part of a problem, indicates that no solution is currently available.*

PROBLEMS

10417. *Proposed by Charles Vanden Eynden, Illinois State University, Normal, IL.*

Characterize the positive integers m such that

$$m^n \equiv 1 \pmod{n} \implies m \equiv 1 \pmod{n}.$$

10418. *Proposed by Răzvan Satnoianu, A. S. E., Bucharest, Romania.*

Given the acute triangle ABC , let h_a , h_b , and h_c denote the altitudes and s the semiperimeter. Show that

$$\sqrt{3} \max \{h_a, h_b, h_c\} \geq s.$$

10419. Proposed by Bill Correll, Jr. (student), Denison University, Granville, OH.

Let k be an integer greater than or equal to 3. Let $S(k)$ be the set of nonnegative real numbers x for which

$$\left\lfloor \frac{x+k-2}{k} \right\rfloor \left\lfloor \frac{x+k-1}{k-1} \right\rfloor + \left\lfloor \frac{x}{k} \right\rfloor = \left\lfloor \frac{x+k-2}{k-1} \right\rfloor \left\lfloor \frac{x+k-1}{k} \right\rfloor + \left\lfloor \frac{x}{k-1} \right\rfloor.$$

(a) Determine the largest integer in $S(k)$.

(b) Show that $S(k)$ is the union of a finite number of intervals with the sum of the lengths of those intervals equal to $(k^2 - 3k + 6)/2$.

10420. Proposed by C. R. Selvaraj and S. Selvaraj, Penn State University — Shenango, Sharon, PA.

Let

$$g_i(n) = \sum_{k=i}^{\infty} \frac{k-i+1}{k!} ((n+2)^k - 2e(n+1)^k + e^2 n^k).$$

Prove that, for all $i > 1$, $g_i(n)$ is a polynomial in n of degree $i - 2$ and $g_i(n) \geq 0$ for all $n \in \mathbb{N}$ and $i \in \mathbb{N}$.

10421. Proposed by Gigel Militaru, University of Bucharest, Bucharest, Romania.

Let n be an integer, $n \geq 3$, and let z_1, \dots, z_n and t_1, \dots, t_n be complex numbers. Prove that there exists an integer i , $1 \leq i \leq n$ with

$$4|z_i t_i| \leq \sum_{j=1}^n |z_i t_j + z_j t_i|.$$

10422. Proposed by Adam Fieldsteel, Wesleyan University, Middletown, CT.

Let $f : [0, 1] \rightarrow \mathbb{R}$ be a C^1 strictly increasing function with $f(1) = L$, where L is the length of the graph of f .

(a) Show that $\int_0^1 f(x) dx \geq \pi/4$.

(b) Show that $\int_0^1 f(x) dx = \pi/4$ only if the graph of f is a quarter circle.

10423. Proposed by M. Filaseta & C. Nicol, University of South Carolina, Columbia, SC.

For a positive integer n , let

$$P_n(x) = \sum \left\{ x^{j-1} : 1 \leq j \leq n, \gcd(j, n) = 1 \right\}.$$

For example: $P_1(x) = P_2(x) = 1$; $P_3(x) = x + 1$; $P_4(x) = x^2 + 1$; $P_5(x) = x^3 + x^2 + x + 1$; and $P_6(x) = x^4 + 1$. Prove that $P_n(x)$ is reducible over the rationals for every $n \geq 7$.

SOLUTIONS

Asymptotic Formulas from Chebyshev Polynomials

6332 [1981, 150]. *Proposed by Petter E. Bjørstad, Stanford University and Henry E. Fettis, Mountain View, CA.*

In analyzing an accelerated numerical integration method for the biharmonic equation, the finite sum

$$S_N = \sum_{k=1}^{N-1} \sin^2 \left(\frac{k\pi}{N} \right) \bigg/ \left[1 + a^2 - 2a \cos \left(\frac{k\pi}{N} \right) \right]^2$$

arises ($0 \leq a^2 < 1$). Find a closed algebraic expression for S_N that exhibits its asymptotic behavior in terms of a and N .

Solution by Heinz-Jürgen Seiffert, Berlin, Germany. We shall prove that

$$S_N = \frac{N}{2(1-a^{2N})} \left[\frac{1+a^{2N-2}}{1-a^2} - 2N \frac{a^{2N-2}}{1-a^{2N}} \right]. \quad (*)$$

Consider the Chebyshev polynomials of the first and the second kind defined by

$$\begin{aligned} T_0(x) &= 1, & T_1(x) &= x, & T_{n+1}(x) &= 2xT_n(x) - T_{n-1}(x) \quad \text{for } n \geq 1, \\ U_0(x) &= 1, & U_1(x) &= 2x, & U_{n+1}(x) &= 2xU_n(x) - U_{n-1}(x) \quad \text{for } n \geq 1, \end{aligned}$$

respectively. It is well-known (I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, prepared by A. Jeffrey, Academic Press, 1980, p. 1032) that

$$T_n(x) = \frac{1}{2} \left[\left(x + \sqrt{x^2 - 1} \right)^n + \left(x - \sqrt{x^2 - 1} \right)^n \right], \quad (1)$$

$$U_n(x) = \frac{\left(x + \sqrt{x^2 - 1} \right)^{n+1} - \left(x - \sqrt{x^2 - 1} \right)^{n+1}}{2\sqrt{x^2 - 1}}, \quad (2)$$

and

$$U_n(\cos \phi) = \frac{\sin((n+1)\phi)}{\sin \phi}. \quad (3)$$

In particular, (3) shows that U_{N-1} has the $N-1$ distinct roots $\cos(k\pi/N)$, $k = 1, \dots, N-1$. Since U_{N-1} has leading coefficient 2^{N-1} we have

$$\prod_{k=1}^{N-1} \left(x - \cos \left(\frac{k\pi}{N} \right) \right) = 2^{1-N} U_{N-1}(x). \quad (4)$$

Differentiating (4) logarithmically gives

$$\sum_{k=1}^{N-1} \frac{1}{x - \cos \left(\frac{k\pi}{N} \right)} = \frac{U'_{N-1}(x)}{U_{N-1}(x)}.$$

Differentiating the latter equation we get

$$\sum_{k=1}^{N-1} \frac{1}{\left(x - \cos \left(\frac{k\pi}{N} \right) \right)^2} = \frac{\left(U'_{N-1}(x) \right)^2 - U_{N-1}(x) U''_{N-1}(x)}{U_{N-1}(x)^2}. \quad (5)$$

Replacing x by $1/x$ in (4) we find

$$\prod_{k=1}^{N-1} \left(1 - x \cos \left(\frac{k\pi}{N} \right) \right) = \left(\frac{x}{2} \right)^{N-1} U_{N-1} \left(\frac{1}{x} \right).$$

The same procedure as above produces

$$\sum_{k=1}^{N-1} \frac{\cos^2 \left(\frac{k\pi}{N} \right)}{\left(1 - x \cos \left(\frac{k\pi}{N} \right) \right)^2} = \frac{N-1}{x^2} + \frac{\left(U'_{N-1} \left(\frac{1}{x} \right) \right)^2 - 2x U_{N-1} \left(\frac{1}{x} \right) U'_{N-1} \left(\frac{1}{x} \right) - U_{N-1} \left(\frac{1}{x} \right) U''_{N-1} \left(\frac{1}{x} \right)}{x^4 U_{N-1} \left(\frac{1}{x} \right)^2}. \quad (6)$$

Now, in (6) we replace x by $1/x$, divide the resulting equation by x^2 , and subtract this result from (5). Since $1 - \cos^2(k\pi/N) = \sin^2(k\pi/N)$, we obtain

$$S_N(x) := \sum_{k=1}^{N-1} \frac{\sin^2 \left(\frac{k\pi}{N} \right)}{\left(x - \cos \left(\frac{k\pi}{N} \right) \right)^2} = \frac{A_{N-1}(x)}{U_{N-1}(x)^2} - N + 1, \quad (7)$$

where

$$A_{N-1}(x) = (x^2 - 1)U_{N-1}(x)U''_{N-1}(x) + 2xU_{N-1}(x)U'_{N-1}(x) - (x^2 - 1)(U'_{N-1}(x))^2. \quad (8)$$

It is well-known (F. G. Tricomi, *Vorlesungen über Orthogonalreihen*, Springer-Verlag, 1955, p.188) that

$$(x^2 - 1)U'_{N-1}(x) = (N-1)xU_{N-1}(x) - NU_{N-2}(x), \quad (9)$$

and

$$(x^2 - 1)U''_{N-1}(x) = (N^2 - 1)U_{N-1}(x) - 3xU'_{N-1}(x). \quad (10)$$

Combining (8), (9) and (10) with a tedious calculation yields

$$(x^2 - 1)A_{N-1}(x) = (N-1)(x^2 - N-1)U_{N-1}(x)^2 + N(2N-1)xU_{N-1}(x)U_{N-2}(x) - N^2U_{N-2}(x)^2. \quad (11)$$

Using

$$xU_{N-1}(x) = (U_N(x) + U_{N-2}(x)) / 2$$

and

$$U_N(x)U_{N-2}(x) = U_{N-1}(x)^2 - 1,$$

(11) becomes

$$(x^2 - 1)A_{N-1}(x) = \left((N-1)(x^2 - 1) + \frac{N}{2} \right) U_{N-1}(x)^2 - \frac{N}{2} U_{N-2}(x)^2 - \frac{N(2N-1)}{2}.$$

Substituting the latter equation in (7) gives

$$S_N(x) = \frac{N}{2} \frac{U_{N-1}(x)^2 - U_{N-2}(x)^2 - 2N + 1}{(x^2 - 1)U_{N-1}(x)^2},$$

or, recalling that

$$2(x^2 - 1)U_{N-1}(x)^2 = T_{2N}(x) - 1$$

and

$$U_{N-1}(x)^2 - U_{N-2}(x)^2 = U_{2N-2}(x),$$

we finally obtain

$$S_N(x) = N \frac{U_{2N-2}(x) - 2N + 1}{T_{2N}(x) - 1}. \quad (12)$$

A straightforward calculation shows that (*) is true for $a = 0$. Since $\cos(\pi - x) = -\cos x$, it is easily seen that the value of S_N does not change if a is replaced by $-a$. Hence, it suffices to prove (*) for $0 < a < 1$. In (12) we take $x = (1 + a^2)/(2a)$. Then we have $S_N(x) = 4a^2 S_N$. From (1) and (2) it follows that

$$T_{2N}(x) = \frac{1}{2}(a^{-2N} + a^{2N}) \quad \text{and} \quad U_{2N-2}(x) = \frac{a}{1-a^2}(a^{-(2N-1)} - a^{2N-1}).$$

Now, (*) easily follows from (12).

It should be noted that, except for $x \in \{\cos(k\pi/N) : k = 1, \dots, N-1\}$, (12) is valid for all complex numbers. Taking $x = \cos \phi$ in (12), using (3) and $T_{2N}(\cos \phi) = \cos(2N\phi)$, a simple calculation yields

$$\sum_{k=1}^{N-1} \frac{\sin^2\left(\frac{k\pi}{N}\right)}{\left(\cos \phi - \cos\left(\frac{k\pi}{N}\right)\right)^2} = N \left(N \csc^2(N\phi) - \cot(N\phi) \cot \phi - 1 \right),$$

valid for all ϕ with $\cos \phi \notin \{\cos(k\pi/N) : k = 1, \dots, N-1\}$. In particular, if N is odd, we may take $\phi = \pi/2$ to get

$$\sum_{k=1}^{N-1} \tan^2\left(\frac{k\pi}{N}\right) = N(N-1) \quad (N = 3, 5, 7, \dots).$$

There is no record of any other solution being received.

Everywhere Unimodal

10243 [1992, 675]. *Proposed by Michel Balazard, Université Bordeaux I, Talence, France.*

Define a sequence of functions $f_k(t)$ for $t > k$ recursively by

$$f_1(t) = 1$$

$$f_{k+1}(t) = \int_k^{t-1} f_k(u) \frac{du}{u}$$

Prove that, for every real number $t > 1$, the sequence $\langle f_k(t) : 1 \leq k < t \rangle$ is unimodal.

Solution by Robin J. Chapman, University of Exeter, U. K. It is clear that $f_k(t) > 0$ for all $t > k$. We claim that for each $k \geq 1$ there exists t_k in $(k+1, \infty]$ such that $f_k(t) > f_{k+1}(t)$ if $k+1 < t < t_k$ and $f_k(t) < f_{k+1}(t)$ if $t > t_k$. To establish this by induction on k , first observe that $f_2(t) = \log(t-1)$, so $t_1 = 1+e$. Then assume that $k > 1$ and that the inductive hypothesis holds for $k-1$. Consider the function g_k defined by $g_k(t) = f_k(t) - f_{k+1}(t)$. We have

$$g_k(t) = \int_{k-1}^k f_{k-1}(u) \frac{du}{u} + \int_k^{t-1} [f_{k-1}(u) - f_k(u)] \frac{du}{u}.$$

The function $g_k(t)$ is increasing for $k+1 < t < t_{k-1}+1$ and decreasing when $t > t_{k-1}+1$. Since it is positive for $k+1 < t < t_{k-1}+1$, it has at most one zero, at a value $t_k > t_{k-1}+1$. (If g_k has no zero, then we put $t_k = \infty$ and $t_\ell = \infty$ for all $\ell \geq k$.) This proves the claim and also that $t_{k+1} \geq t_k + 1$ for all k .

To show unimodality it suffices to show "once a decrease always a decrease." Consider a fixed t . If $f_k(t) \geq f_{k+1}(t)$, then $t \leq t_k$, and hence $t < t_{k+1}$. Thus $f_{k+1}(t) > f_{k+2}(t)$, and the result follows.

Editorial comment. The proposer's solution worked with the function r_k defined by $r_k(t) = f_{k+1}(t)/f_k(t)$, and showed that $r_k(t)$ is strictly increasing in t for $t > k + 1$. This led to $f_{k+1}^2(t) > f_k(t)f_{k+2}(t)$ for $t > k + 2$, hence the sequence $f_k(t)$ is logarithmically concave for $1 \leq k < t$.

The proposer introduced these functions to study the unimodality of the distribution of the number of prime divisors of an integer. In fact, let $F(x)$ be the number of positive integers n not exceeding x that have exactly k prime divisors and such that each of these exceeds $x^{1/t}$. Then

$$f_k(t) = \lim_{x \rightarrow \infty} F(x) / (x / \log x).$$

For $k = 1$ and $t > 1$ this is the prime number theorem.

He also remarks that if $k_0(t)$ is any function such that $r_k(t) > 1$ for $k < k_0(t)$ and $r_k(t) < 1$ for $k > k_0(t)$, then there is an elementary proof that $k_0(t) \leq \log t + 1$. The theory outlined here leads to a lower bound on $k_0(t)$ that is also of the form $\log t + O(1)$ but he claims that, "the proof is no longer elementary".

Solved also by the proposer.

An Odd Square

10263 [1992, 873]. *Proposed by J. G. Mauldon, Amherst College, Amherst, MA.*

Let m and n be odd integers, and suppose that $n^2 - 1$ is a multiple of $|m^2 + 1 - n^2|$. Prove or disprove that this requires that $|m^2 + 1 - n^2|$ be the square of an integer.

Solution by the proposer. The implication holds, and it also holds if m, n are nonzero even integers. Furthermore, $m^2 \geq n^2$, so that $m^2 + 1 - n^2$ itself is a square.

By hypothesis, we have $n^2 - 1 = t(m^2 + 1 - n^2)$ for some integer t , which we rewrite as $m^2 = (t + 1)(m^2 + 1 - n^2)$. Let $k = t + 1$; this integer is nonzero. It suffices to show that k is a square, since that will show that the integer $m^2 + 1 - n^2$ is a rational square. However, the only integers that are squares of rationals are squares of integers.

Let S_k be the set of integer ordered pairs (x, y) such that $(x + y)^2 = k(1 + 4xy)$. Because $((m+n)/2, (m-n)/2) \in S_k$, the set is nonempty. Let $a = \min\{|x| : (x, y) \in S_k\}$. It suffices to show that $a = 0$.

The set S_k is invariant under negation and under interchange of coordinates. Hence a belongs to a pair in S_k . The other member of such a pair must solve the quadratic equation $(x + a)^2 = k(1 + 4ax)$. Hence the two roots b_1, b_2 must both be integers and have absolute value at least a . From the quadratic equation, we have $b_1 + b_2 = 4ak - 2a$ and $b_1 b_2 = a^2 - k$. By direct computation, these yield $(a + b_1)(a + b_2) = (4a^2 - 1)k$.

If $k < 0$, then the equations above yield $b_1 b_2 > 0$ and $b_1 + b_2 \leq 0$. Hence each of b_1, b_2 is negative. Since $a < |b_i|$, this implies $a + b_i < 0$, and hence $(4a^2 - 1)k > 0$. With $k < 0$, this requires $a = 0$, but we have noted that $a = 0$ implies k is a square. Hence we may assume $k > 0$.

If $k > 0$ and $a > 0$, the equations yield $b_1 + b_2 > 0$ and $(a + b_1)(a + b_2) > 0$, so each of b_1, b_2 is positive. By the choice of a , we have $a^2 \leq b_1 b_2$, but also $k > 0$ implies $b_1 b_2 = a^2 - k < a^2$. The contradiction implies $a = 0$, and hence k is a square.

Editorial comment. The other correct solutions reduced the problem to solving a Pell equation (i. e., a Diophantine equation of the form $u^2 - Dv^2 = 1$ for some non-square positive integer D). In this case, D is the *squarefree part* of $k(k - 1)$, and the special form of D is exploited in the solution. The incorrect solution claimed that the conditions hold only when $m = \pm n$ or $n = \pm 1$, which omits the solution $(m, n) = (105, 99)$.

Solved also by J. C. Binz (Switzerland), R. J. Chapman (U. K.), I. Kastanas, J. P. Robertson, and the GCHQ Problem Solving Group (U. K.). One incorrect solution was received.

Polygons with Inscribed Circles

10303 [1993, 401]. *Proposed by David E. Gurarie, Case Western Reserve University, Cleveland, OH.*

Let a_1, \dots, a_n ($n \geq 3$) be positive real numbers.

(a) Find necessary and sufficient conditions on a_1, \dots, a_n for there to exist a convex n -gon which admits an inscribed circle and whose sides, in cyclic order, are a_1, \dots, a_n .

(b) Find the radius of the inscribed circle.

Solution by Richard Holzsager, The American University, Washington, DC. Suppose that $V_0, V_1, \dots, V_n = V_0$ are the vertices of a convex polygon with an inscribed circle, and with side $V_{j-1}V_j$ of length a_j , $j = 1, \dots, n$. Let d_j be the distance from V_{j-1} to the point of tangency of the inscribed circle on $V_{j-1}V_j$. Then

$$a_j = d_j + d_{j+1} \quad (1)$$

(with $d_{n+1} = d_1$), and

$$d_j > 0. \quad (2)$$

The existence of such a set of numbers d_j is also sufficient for the existence of an inscribed circle. In fact, consider the function $f(r) = \sum \arctan(d_j/r)$, which is monotonically decreasing for $r \geq 0$, from $n\pi/2$ down to 0. If we choose the unique r with $f(r) = \pi$, and form the polygon circumscribed about a circle of radius r , tangent at points around the circle at successive angles $2\arctan(d_j/r)$, then the successive sides are the required a_j . It remains to give criteria for the existence and positivity of the d_j .

If the d_j exist, then the alternating sum $\sum (-1)^{j-1} a_j$ collapses: to 0 if n is even; to $2d_1$ if n is odd. In the odd case, d_1 , and similarly all the d_j 's, are determined from the a_j . So the solution, if it exists, is unique in this case. Furthermore, these alternating sums lead to expressions for d_j that clearly satisfy (1). Positivity of these alternating sums is therefore necessary and sufficient for the existence of the circle.

In the even case, we can choose any d_1 , and satisfy (1) by setting $d_{j+1} = a_j - d_j$ for $j = 1$ to $n-1$. The successive requirements for positivity can then be stated $d_1 > 0$, $d_1 < a_1$, $d_1 > a_1 - a_2$, $d_1 < a_1 - a_2 + a_3$, etc. A necessary and sufficient condition is therefore that the maximum of 0, $a_1 - a_2$, $a_1 - a_2 + a_3 - a_4, \dots$ be less than the minimum of a_1 , $a_1 - a_2 + a_3, \dots$.

In a sense we already have the answer to part b, namely $r = f^{-1}(\pi)$, where $f(r) = \sum \arctan(d_j/r)$. To make this a bit more tractable, note that $f(r)$ is an integral multiple of π iff $\Im(\prod (r + d_j i)) = 0$, or $\sigma_1 r^{n-1} - \sigma_3 r^{n-3} + \sigma_5 r^{n-5} \dots = 0$, where the σ_j are the symmetric functions of the d_j . To get $f(r) = \pi$, i. e., the smallest positive integral multiple of π , the fact that f is decreasing shows that we need the largest root of the equation.

Also note that there is an obvious relation $A = r \sum a_j/2$ between r and the area A of the polygon, shown by cutting up into triangles $V_{j-1}V_jC$, where C is the center of the circle. For a triangle, this relation, joined with the solution $r = \sqrt{\sigma_3/\sigma_1}$, gives Heron's formula for the area in terms of the sides.

Editorial comment. Several solvers noted that a rhombus of side 2 allows $0 < r \leq 1$. The strict inequality $0 < r$ is a consequence of the strict inequality in (2). This suggests investigating the possible degenerate cases introduced by requiring only that the d_j be nonnegative. Clearly, $d_{j+1} = 0$ corresponds to both $V_{j-1}V_j$ and V_jV_{j+1} lying on the tangent to the circle at V_j . The angle at V_j then degenerates to a *straight angle* and the resulting polygon has fewer than n sides. On the other hand, Ilias Kastnas noted that the case of $n = 2k$ can be viewed as a degenerate case of $n = 2k + 1$. The choice of d_1 amounts to locating an additional vertex on V_0V_1 .

For $n = 4$, the algebraic expression for r as a function of d_1 is easily analyzed. However, as in the case $n = 3$, r is proportional to area when the side lengths are given. The study of possible values of r is thus a special case of the problem of finding extremes of area for convex quadrilaterals with given edge lengths.

Solved also by R. Barbara (Lebanon), V. Božin (student, Yugoslavia), R. J. Chapman (U. K.), B. N. Cheng (The Philippines), S. M. Gagola Jr., I. Kastanas, J. H. Lindsey II, O. P. Lossers (The Netherlands), H. Morris, A. Nijenhuis, R. M. Robinson, R. A. Simon (Chile), GCHQ Problem Solving Group (U. K.), and the proposer.

REVIVALS

A Permutation on the Cube

6670 [1991, 862; 1993, 595]. *Proposed by R. H. Jeurissen, Toernooiveld, Nijmegen, The Netherlands.*

Let $\{0, 1\}^n$ denote the set of n -bit strings of zeros and ones. If $(a_1, \dots, a_n) \in \{0, 1\}^n$, let $\pi_n(a_1, \dots, a_n)$ be the string (b_1, \dots, b_n) given by $b_1 = a_1$ and $b_k \equiv a_k + a_{k-1} \pmod{2}$ for $1 < k \leq n$. Since (a_1, \dots, a_n) can be retrieved from (b_1, \dots, b_n) , it is clear that π_n is a permutation of $\{0, 1\}^n$. Determine the cycle structure of the permutation π_n , i.e., the lengths of the cycles that occur and the number of cycles of each length.

Editorial comment. David Singmaster has noted that π_n is a famous permutation that may not be recognized here because of the emphasis on its cycle structure. If we let $B_n(k)$ denote the n place binary representation of k for $0 \leq k < 2^n$, then $G_n(k) = \pi_n(B_n(k)) = B_n(k) \oplus B_n(\lfloor k/2 \rfloor)$, where \oplus denotes the vector space addition in $\{0, 1\}^n$. The sequence $G_n(0), G_n(1), \dots, G_n(2^n - 1)$ is known as the Gray code. It gives a Hamiltonian circuit on the n -cube.

Adjacent strings in the Gray code also give the possible moves in the *Chinese Rings* puzzle. To solve the puzzle, one must find k from $G_n(k)$. Since k is easily recovered from $B_n(k)$, it suffices to have a formula to invert π_n . Observations made in the published solution give the formula $B_n(k) = G_n(k) \oplus G_n(\lfloor k/2 \rfloor) \oplus \dots \oplus G_n(\lfloor k/2^{n-1} \rfloor)$. This method of analyzing the puzzle is attributed to Louis A. Gros who published a pamphlet on the puzzle in 1872.

Products of Nilpotent Matrices

10200 [1992, 163; 1993, 807]. *Proposed by Daniel Goffinet, St. Étienne, France.*

(a) Prove that a (square) matrix over a field F is singular if and only if it is a product of nilpotent matrices.

(b) If $F = \mathbb{C}$, prove that the number of nilpotent factors can be bounded independently of the size of the matrix.

Editorial comment. It has been shown over an arbitrary field that: for M a singular matrix that is not a 2 by 2 nonzero nilpotent matrix, $M = AB$ with A and B nilpotent. This result was known over the field of complex numbers, but the best result over a general field submitted as a solution of this problem led to a product of four factors. Pei Yuan Wu has submitted references to proofs of the general result.

In T. J. Laffey, "Products of matrices" in *Generators and Relations in Groups and Geometries* (A. Barlotti, E. W. Ellers, P. Plaumann & K. Strambach, eds.), Kluwer, Dordrecht,

1991, 95–123, the space on which M acts is written as a direct sum of invariant subspaces $V_0 \oplus V_1$ with the action of M being nilpotent on V_0 and nonsingular on V_1 . Using known factorizations of nilpotent matrices, and of nonsingular matrices, a factorization of a matrix similar to M is given. Several cases must be considered to give a complete construction.

In A. R. Sourour, “Nilpotent factorizations of matrices”, *Linear Multilinear Algebra* 31 (1992), 303–308, an inductive proof is used. The key is a general proposition on partitioning matrices over a field F .

Proposition. *Let $A \in M_n(F)$. Then A is similar to a matrix of the form*

$$\begin{bmatrix} \alpha & c^t \\ b & D \end{bmatrix}$$

with $\alpha \in F$, $D \in M_{n-1}(F)$, $\text{rank } D = \text{rank } A - 1$, $b \in \text{Range } D$ and $c \in \text{Range } D^t$ if and only if $A^2 \neq 0$.

This proposition, combined with a study of matrices with $M^2 = 0$ achieves a factorization into nilpotent matrices each of which has the same rank as M .

Collaborating editors: David F. Appleyard, Paul T. Bateman, Bruce C. Berndt, Duane M. Broline, Barry W. Brunson, Frank S. Cater, Gulbank D. Chakerian, Underwood Dudley, Gerald A. Edgar, Michael A. Filaseta, Ira M. Gessel, Richard A. Gibbs, Jerrold R. Griggs, Douglas A. Hensley, John R. Isbell, Mourad E. H. Ismail, Murray Klamkin, Daniel J. Kleitman, Frederick W. Luttman, Frank B. Miles, Richard Pfeifer, Stephen L. Portnoy, J. O. Shallit, John Henry Steelman, Kenneth B. Stolarsky, David E. Tepper, Douglas B. Tyler, Daniel Ullman, and William E. Watkins.

In addition to the Collaborating editors, the following individuals served as Referees or Guest Editors for material considered for publication in 1994: George E. Andrews, Richard A. Askey, George Baloglou, Gilbert Baumslag, Jozsef Beck, Grahame Bennett, Peter B. Borwein, Ezra A. Brown, David G. Cantor, Bille Carlson, Sagun Chanillo, Gregory L. Cherlin, J. Brian Conrey, John H. Conway, David A. Cox, Dennis DeTurck, Persi Diaconis, Vladimir Drobot, Harold M. Edwards, Noam Elkies, Edgar A. Feldman, Joseph A. Gallian, Fred Galvin, Murray Gerstenhaber, Bart Goddard, Gene H. Golub, Richard F. Gundy, Richard K. Guy, Aimo Hinkkanen, Roger Horn, James E. Humphreys, Geoffrey A. Kandall, J. H. B. Kemperman, Clark Kimberling, Victor Klee, Solomon Leader, James I. Lepowsky, Eugene M. Luks, Feng Luo, Victor S. Miller, Howard Morris, Benjamin Muckenhoupt, Melvyn B. Nathanson, Roger D. Nussbaum, Ingram Olkin, M. J. Pelling, Carl Pomerance, Stanley Rabinowitz, Mizanur Rahman, M. Rychlik, Michael Saks, Lawrence A. Shepp, Richard P. Stanley, Kenneth B. Stolarsky, Gilbert Strang, Howard M. Taylor, Charles H. Toll, Jerrold B. Tunnell, Charles L. Vanden Eyn-den, Wolmer V. Vasconcelos, Bertram Walsh, Edward T. H. Wang, William C. Waterhouse, Gregory P. Wene, Richard Wheeden, Herbert S. Wilf, Eduard Wirsing, and Doron Zeilberger. The editors wish to thank them for their help in preparing this column.

REVIEWS

Edited by **Darrell Haile**
Indiana University, Bloomington IN 47405

Linear Programs and Related Problems, by Evar D. Nering and Albert W. Tucker, Academic Press, San Diego, 1993; 578 + pp

Reviewed by **Stephen B. Maurer**

This undergraduate text is a joint effort by the author of several fine texts (Nering) and one of the grand old men of mathematics and a former MAA President (Tucker). It should be on the shelf of anyone who teaches linear programming, and it invites careful consideration for classroom adoption. It provides an approach that deserves to be better known, for it combines theory and practice and makes key ideas (notably duality) particularly transparent. Most faculty will learn at least as much from the book as students will.

Linear programming (LP) is the study of optimizing linear functions subject to linear inequality constraints. The subject is blessed to have both beautiful theory and myriad applications. Consequently, there are many LP texts of several varieties. Most common are books with an Operations Research orientation, such as Hillier and Lieberman [6] and Bradley, Hax and Magnanti [2]. These books regard LP as a tool for modeling. The theory is usually there, but as a sideshow. They emphasize how to recognize and interpret a linear program in a real-world situation and how to apply algorithmic implementations that are efficient and numerically stable.

In contrast, books more mathematical in flavor tend to emphasize the theory and structure of the subject. They often dwell on the connection to n -dimensional polyhedra, and either downplay algorithms or highlight the more theoretical algorithmic issues from complexity theory. Books written in this vein are typically at the graduate level. A fine recent example is Schrijver [12]. Older and less advanced is Gass [5].

There are also LP books written by economists, which emphasize the economic applications and interpretations. Two classics are Dorfman, Samuelson and Solow [4] and Baumol [1].

This division of texts recapitulates history. When George Dantzig, at the Pentagon after World War II, was first able to secure funding to promote this new subject, three groups got started: one under Tucker at Princeton to develop the theory, another under Koopmans at Chicago to explore connections with economics, and the third under Dantzig to explore algorithms, specifically, the simplex algorithm. For more on the history, see [8, 9, 13].

Note the schism, deliberately introduced by the players themselves, between theory and practice. It doesn't have to be that way. LP theory can be developed from the algorithms, and you can delight in both from the start. A key strength of this book is that it shows how. For instance, suppose an algorithm to compute the

optimum of a function f has the properties that

- it has only a finite number of states
- it does not cycle among states
- if its current state neither provides an optimal value of f nor shows that f is unbounded, then the algorithm passes to another state.

Then it is a mere observation that f must either attain an optimum or be unbounded. Yet it is just this observation that one may use with the simplex algorithm to show that any feasible linear program either attains an optimum or is unbounded. (Feasible means the domain of f is nonempty.)

As the authors put it (p. 126)

Many people think of an algorithm as a method for finding a solution to a problem for which one already knows, by some other means, that a solution exists. We look on algorithms as much more than that and we use them for much more than that. In many cases an algorithm can supply the proof that the desired result exists. . . .

In other words, Nering and Tucker are advocates of *proof by algorithm*, a refinement of the old idea of constructive proof.

Today, proof by algorithm is a well-known concept. Certainly in the field of combinatorial optimization it is the preferred method of proof—in this field existence proofs are not considered more esthetic. But since this preference is a rather drastic change from the mathematical esthetic of mid century, one can ask how the change came about. A reasonable assumption is that it came from Computer Science. However, it may be that Tucker had a lot to do with it. He started turning to the algorithmic viewpoint in the mid 1950s (see the interview with him [10]), and his group at Princeton included or interacted with almost all the early workers in combinatorial optimization. There is a fine math history research project waiting here—how *did* this new esthetic come about?

A second great strength of this book is its treatment of duality. Linear programs come in pairs. To take the canonical case, if the “primal” problem is to maximize cx subject to $Ax \leq b$ and $x \geq 0$, then the dual is to minimize vb subject to $vA \geq c$ and $v \geq 0$. (Lowercase letters are vectors, uppercase matrices.) Dual problems interact. For instance, if both problems are feasible, then it turns out both attain optima and the optimum values are equal. Duality has many important consequences, e.g., a lazy supervisor test for checking a claimed optimum, alternative algorithms, shadow prices in economics.

All LP texts cover duality, but usually as a somewhat mysterious add-on. This too mimics history, but there is a better way. Dual programs can be introduced simultaneously by the use of a special representation, the *condensed* or *Tucker tableau*:

$$\begin{array}{cccccc}
 & x_1 & x_2 & \cdots & x_n & -1 \\
 v_1 & a_{11} & a_{12} & \cdots & a_{1n} & b_1 & = -y_1 \\
 v_2 & a_{21} & a_{22} & \cdots & a_{2n} & b_2 & = -y_2 \\
 \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\
 v_m & a_{m1} & a_{m2} & \cdots & a_{mn} & b_m & = -y_m \\
 -1 & c_1 & c_2 & \cdots & c_n & d & = f \\
 & = u_1 & = u_2 & \cdots & = u_n & = g &
 \end{array} \tag{1}$$

The variables on the top multiply the columns to obtain the variables on the right, i.e., $Ax + (-1)b = -y$, or equivalently, $Ax + y = b$, or $Ax \leq b$ if all variables are nonnegative. These row equations are the constraints, and the basement row equation, $cx - d = f$, is the equation for the “objective” function f to be maximized. This is the primal problem. The variables on the left and bottom give the dual: minimize $vb - d = g$ subject to $vA - c = u$, or equivalently, $vA \geq c$.

In brief, to develop the theory from (1), one shows by elementary linear algebra that the simplex algorithm just exchanges variables on the top and right, so that different variables become “basic” when the variables on the top are set to 0. Of course, the entries in the tableau must be updated too, so that the equal signs on the right are still correct after variables are exchanged. Marvelously, the same algebra maintains the equal signs on the bottom as the dual variables are exchanged. The conditions that make the basic solution of a tableau optimal are simple to understand and describe—the entries in b must be nonnegative and those in c nonpositive. By (skew) symmetry these same conditions describe optimality for the dual. Thus a condensed tableau can exhibit optimality for both problems simultaneously and, since x and v are 0 for a basic solution, the objective functions f and g have the same value, d .

Almost all LP books use tableaus, but most use some version of extended form, where $A = [a_{ij}]$ in (1) gets replaced by $[A \ I]$ and the “slack” variables y in (1) move to the columns of I . But it is much harder to even see a dual, let alone develop its theory, in these formats. A few books do use condensed tableaus, but without the variables on the left and bottom, which is the key to handling duality. To my knowledge, the only other texts with true condensed tableaus are Kemeny Snell and Thompson’s “Finite Mathematics”, third edition only [7], and Rothenberg [11].

Part I of Nering and Tucker develops linear programming itself, along the lines described above. Part II treats various related problems—related in that most can be stated as LPs but their special form allows for special algorithms. The emphasis on proof by algorithm continues, and duality is often used to clarify the algorithms (but not as often as it could be). For instance, the minimax theorem of matrix games is reduced to the existence-duality theorem of LP. Kuhn’s Hungarian method for the minimum weight assignment problem is shown to work by increasing the sum of the dual variables until the primal and dual objective functions are equal. The same approach is used for analyzing Dantzig’s algorithm for the transportation problem. Other topics include various network problems—transshipment, maximum flow, shortest path—and an introduction to nonlinear programming—the Karush-Kuhn-Tucker Theorem (still known as the Kuhn-Tucker theorem to many) and various quadratic programs.

Tucker once said that his goal in mathematics has always been to “unify and simplify”. The fruits of this attitude appear in this book. It’s a shame that his condensed tableau format is not better known.

These days most talk in LP circles is about completely different approaches: Khachian’s 1979 ellipsoid algorithm, Karmarkar’s 1984 interior point method, and more recent variations. Khachian’s algorithm was important because it is polynomial: there is a polynomial P so that his algorithm will solve every LP problem in $P(n)$ steps, where n is the amount of input data. However, in practice the Ellipsoid algorithm is much slower than simplex. (The simplex algorithm takes linear time on average but special cases take exponential time.) Karmarkar’s algorithm is not only polynomial, but fast in practice on at least some sorts of problems. The jury is still out on what is the best commercial method.

Nering and Tucker are thin, and mostly nontechnical, on the new methods. For Khachian's algorithm, no algebraic description or analysis is given, nor are there numerical examples or homework problems. For Karmarkar's algorithm, some algebraic detail is given, one iteration of one problem is shown, and there is one homework problem. (This material appears in the final chapter of Part I, which also introduces numerical issues and provides ties to other simplex approaches.)

This thinness is understandable. There are no small, complete examples for the new methods, and their theory is much more complicated than for the simplex method. Also, no one has shown a nice way to develop the existence and duality theory of LP from these algorithms. Since building the theory from algorithms is a key theme of this book, the authors rightly emphasize simplex.

Nonetheless, since the new methods are at least part of the future, it would be good for students to get more than a passing nod at them even in a first course. Software seems called for. A program with an option to hide the numbers and show pictures might be best. To my knowledge, no text has appeared that takes this approach.

How will students like this book? The authors have made LP theory simple (but still rigorous!) by reducing it to displays and arguments that are carried forward by linear equations and numerical calculations. The writing is clear, straightforward and leisurely, but also somewhat bland. The book is written more in essay style than in textbook chunk style—example, theorem, proof, sidebar, vignette, etc. Much as this chunk style has been criticized, students are used to it and it does allow them clear touchstones and stopping points in what is usually difficult material for them. In Nering and Tucker, important points often appear in the middle of paragraphs in pages full of text.

Based on my experience teaching from a preliminary version some years ago, and more recent reports, I would say that students won't find the book hard to understand—the usual math book complaint—but they won't find it exciting either. It can't be very neat mathematics if it is just about linear equations and a lot of number manipulation, can it? Of course it can, but the teacher's biggest job in using this text will be to explain to students why. Alas, if you make something sufficiently simple for readers who don't know that it used to be complicated, they won't appreciate what they have!

A very nice set of real-world-like examples and problems are introduced in the first chapter: problems too simplified to be really applied, but they show the way. In later chapters, the problems (though not the examples) are mostly purely mathematical. Problems appear at the ends of chapters only. A rather complete set of solutions is supplied in the back, an unusual feature. Also, menu-driven software is provided with the book—for PCs, but a Macintosh version is due by the end of 1993. With this software one can do the arithmetic for all the problems, in the format used by the book, either all at once, or step by step. The data for all the problems is already read in. I found this software serviceable but not effortless. The instructions are terse.

Readers intrigued by this book might also look at Chvátal [3]. It is like-spirited in doing theory, algorithmics and applications simultaneously, for a similar audience. It has a somewhat livelier format and style, and a large variety of problems. It covers many more topics and is more advanced (however, much advanced material is identified with small print and can be skipped). On the other hand, Chvátal does not have Tucker tableaus (or any tableaus), and duality, though done early, is not present from the start.

To conclude, a founding father of mathematical programming, involved in it for 45 years, has put the full maturity of his perspective into this book. His perspective is quite personal, and the excitement and beauty seen by both authors may not come across fully to you or your students. But if you teach LP you impoverish yourself by not taking a look.

Note: Prof. Maurer was a Ph.D. student of Tucker, and learned LP from him.

REFERENCES

1. William J. Baumol, *Economic theory and operations analysis* 4th ed., Prentice-Hall, Englewood Cliffs, N.J., 1977.
2. Stephen P. Bradley, Arnoldo C. Hax and Thomas L. Magnanti, *Applied mathematical programming*, Addison-Wesley, Reading, MA, 1977.
3. Vašek Chvátal, *Linear programming*, W. H. Freeman & Co, New York, 1983.
4. Robert Dorfman, Paul A. Samuelson and Robert M. Solow, *Linear programming and economic analysis*, McGraw-Hill, New York, 1958.
5. Saul I. Gass, *Linear Programming*, 5th ed., McGraw-Hill, New York, 1985.
6. Frederick S. Hillier and Gerald J. Lieberman, *Introduction to Operations Research*, 4th ed., Holden-Day, Inc., Oakland, CA, 1986.
7. John Kemeny, Laurie Snell and Gerald Thompson, *Finite Mathematics*, 3rd ed., Prentice-Hall, Englewood Cliffs, NJ, 1974.
8. Harold W. Kuhn, Nonlinear programming: a historical view, *SIAM-AMS Proceedings*, vol. IX, AMS, Providence, RI, 1976.
9. Jan Karel Lenstra, Alexander Kan and Alexander Schrijver, eds., *History of Mathematical Programming: a Collection of Personal Reminiscences*, North-Holland, Amsterdam, 1991.
10. Stephen B. Maurer, An interview with Albert W. Tucker, *Two-Year College Mathematics Journal*, Vol. 14, No. 3 (June 1983), 210–224.
11. Ronald Rothenberg, *Linear Programming*, North Holland, New York, 1979.
12. Alexander Schrijver, *Theory of linear and integer programming*, Wiley, New York, 1986.
13. E. Roy Weintraub, ed., *Toward a history of game theory*, Duke University Press, Durham, NC, 1992.

Department of Mathematics
Swarthmore College
Swarthmore, PA 19081-1397
smaurer1@cc.swarthmore.edu

The real danger is not that computers will
begin to think like men, but that men will
begin to think like computers.

—*Sydney J. Harris*

Howard W. Eves, *Return to Mathematical Circles*,
Boston: Prindle, Weber and Schmidt, 1988.

TELEGRAPHIC REVIEWS

Edited by **Arnold Ostebee and Paul Zorn**

with the assistance of the Mathematics Departments of
Carleton, Macalester, and St. Olaf Colleges

Telegraphic Reviews are designed to alert readers in a timely manner to new books and computer software appropriate to mathematics teaching and research. Special codes classify reviews by subject area and appropriate use:

<i>T</i> : Textbook	<i>P</i> : Professional Reading	<i>1-4</i> : Semester
<i>C</i> : Computer Software	<i>L</i> : Undergraduate Library	<i>**</i> : Special Emphasis
<i>S</i> : Supplementary Reading	<i>13</i> : Grade Level	<i>??</i> : Questionable

Readers are advised that price information is subject to change. Selected books and software packages receive a second, more extensive review in the *Monthly*.

Books and software submitted for review should be sent to *Book Reviews Editor*, *American Mathematical Monthly*, St. Olaf College, 1520 St. Olaf Avenue, Northfield, MN 55057-1098.

General, L. *Intersections*. JoAnne Growney. Kadet Pr (Bloomsburg, PA 17815-1758), 1993, 56 pp, \$6.95 (P). [ISBN 0-9637964-0-2] A slim book of spare, accessible, sometimes elegant poems. About 1/4 are on mathematical subjects, the rest on other aspects of an interesting, varied life. PZ

General, L. *Helaman Ferguson: Mathematics in Stone and Bronze*. Clàire Ferguson. Meridian Creative Group (5178 Station Rd, Erie, PA 16510), 1994, xv + 79 pp, \$39.95. [ISBN 0-9639121-0-0] Gorgeous photographs, artistic commentaries (by sculptor's wife), and mathematical discussions (by sculptor) of 29 striking works by mathematician/sculptor Helaman Ferguson. Though inspired mathematically, sculptures go beyond illustration, incorporating physical and natural forms and materials. PZ

Mathematics Appreciation, S(13). *Fractal Worlds*. Susan Brendel. J Weston Walch, 1994, iii + 4 pp, \$14.95 poster set (w/black-line masters). [ISBN 0-8251-2464-6] Four 17" x 22" color posters on classical fractals, the Mandelbrot set, Julia sets, and fractal universes. Only the classical fractals poster (Cantor dust, Sierpinski carpet and gasket, Koch island) gives enough information for understanding. MW

Precalculus, T*(13: 1, 2). *Precalculus: Functions and Graphs: A Graphing Approach*. Roland E. Larson, Robert P. Hostetler, Bruce H. Edwards. DC Heath, 1994, xxxii + 973 pp. [ISBN 0-669-35206-3] Visualization emphasis: many graphs and graphical problems; notes on technology use. Includes discussion and

writing problems, sections on exploring data and applications. TH

Education, S(16-18), P*. *Learning Activities from the History of Mathematics*. Frank J. Swetz. J Weston Walch, 1994, viii + 269 pp, \$23.95 (P). [ISBN 0-8251-2264-3] Excellent resource for secondary school teachers. Biographical sketches (with library investigation questions) of 23 mathematicians from Thales to Bourbaki. Student activity sheets explore historical problems. Wealth of quotations and extensive resource list. MW

Education, S(15-17). *Teaching with Student Math Notes, Volume 2*. Ed: Evan M. Maletsky. NCTM, 1993, ix + 123 pp, \$14.50 (P). [ISBN 0-87353-369-0] 20 issues of *Student Math Notes* from 1986-1990 NCTM *News Bulletin*. One theme per 4-page issue. Teacher notes include solutions, extensions. Perforated pages permit classroom duplication. Classroom-tested ideas for secondary teachers. MW

History, S, L. *A History of Vector Analysis: The Evolution of the Idea of a Vectorial System*. Michael J. Crowe. Dover, 1994, xvii + 270 pp, \$7.95 (P). [ISBN 0-486-67910-1] Corrected reprint of 1985 edition (TR, February 1987). LC

History, P. *NBS-INA—The Institute for Numerical Analysis—UCLA 1947-1954: Mathematicians Learning To Use Computers*. Magnus R. Hestenes, John Todd. NIST Special Pub., No. 730. US Government Printing Office (Supt. of Documents, Washington, DC 20402), 1991, ix + 471 pp, \$12.50 (P). The INA was set up to ensure that some mathematicians could handle

the Turing computer. This history stresses research and educational aspects of the INA. **DH History, S(16–18), P, L.** *A History of Complex Dynamics: From Schröder to Fatou and Julia.* Daniel S. Alexander. Aspects of Math., V. E24. Friedr Vieweg & Sohn, 1994, viii + 165 pp, \$42. [ISBN 3-528-06520-6] History of iteration of complex maps from Schröder's 1870 paper (examining Newton's method) through works of Julia, Fatou, and Montel. Good reading for graduate students and mathematicians—specialists and nonspecialists alike. **KS**

History, S(13–16), P, L. *George Green: Mathematician and Physicist 1793–1841: The Background to his Life and Work.* D.M. Cannell. Athlone Pr, 1993, xxvi + 265 pp, \$70. [ISBN 0-485-11433-X] A complete picture of Green's life and education. His contributions to mathematical physics are huge, but relatively little known. A miller, untrained in mathematics until age 40, Green remained a scientific outsider. Author has devoted decades to restoring Green's mill and to studying his life and work. **KS**

History, P*, L. *Möbius and his Band: Mathematics and Astronomy in Nineteenth-century Germany.* Eds: John Fauvel, Raymond Flood, Robin Wilson. Oxford Univ Pr, 1993, 172 pp, \$29.95. [ISBN 0-19-853969-X] In 1810 Germany was a scientific backwater; by 1860 it led the world. In this period Möbius rose from undergraduate to full professor and observatory director at Leipzig. The first five essays describe the flowering of German science, treating Möbius' life and work as archetype. The final essay surveys highlights of modern celestial mechanics/chaotic dynamics. **SK**

Logic, P, L. *Metaphysical Myths, Mathematical Practice: The Ontology and Epistemology of the Exact Sciences.* Jody Azzouni. Cambridge Univ Pr, 1994, ix + 249 pp, \$54.95. [ISBN 0-521-44223-0] Argues against mathematics imitating practices of the empirical sciences. Backbone of mathematical practice is seen as algorithmic development, governed by axioms. Mathematics' primary tool—proof—distinguishes it from other sciences. **RM**

Discrete Mathematics, T*(13), C, L. *Discrete Dynamical Modeling.* James T. Sandefur. Oxford Univ Pr, 1993, xiii + 428 pp, \$39.95. [ISBN 0-19-508438-1] Elementary introduction, suitable for general audiences. Topics include first- and higher-order systems, probabilistic models, and systems. Excellent examples and problems from many areas. **MPR**

Discrete Mathematics, P. *Selected Topics in Discrete Mathematics.* Ed: A.K. Kelmans. Transl., Ser. 2, V. 158. AMS, 1994, xiii +

221 pp, \$79. [ISBN 0-8218-7509-4] 19 papers, on disparate subjects, from the Moscow Discrete Mathematics Seminar (1972–1990).

Number Theory, S(18), P. *Catalan's Conjecture: Are 8 and 9 the Only Consecutive Powers?* Paulo Ribenboim. Academic Pr, 1994, xv + 364 pp, \$64.95. [ISBN 0-12-587170-8] Author recaps results on a variety of Diophantine problems. Elementary, algebraic, and analytic approaches are described. **SG**

Number Theory, P. *Motives.* Eds: Uwe Jannsen, Steven Kleiman, Jean-Pierre Serre. Proc. of Symp. in Pure Math., V. 55, Parts 1 & 2. AMS, 1994, \$250 set [ISBN 0-8218-1635-7]; *Part 1*, xiv + 747 pp, \$140; *Part 2*, xiv + 676 pp, \$129. 47 papers (introductions, specialized surveys, and research reports) from a 1991 AMS-IMS-SIAM Summer Research Conference at the University of Washington.

Number Theory, P. *Décomposition Spectrale et Séries d'Eisenstein: Une Paraphrase de l'Écriture.* Colette Mœglin, Jean-Loup Waldspurger. Prog. in Math., V. 113. Birkhäuser, 1994, xxix + 320 pp, \$98. [ISBN 0-8176-2938-6]

Group Theory, T(17–18: 1, 2), P.** *Computation with Finitely Presented Groups.* Charles C. Sims. Ency. of Math. & Its Applic., V. 48. Cambridge Univ Pr, 1994, xiii + 604 pp, \$99.95. [ISBN 0-521-43213-8] Clear, extensive, thorough, accessible. Treats rewriting systems; automata and rational languages; coset enumeration; the Reidemeister-Schreier procedure; computations with Abelian groups, polycyclic groups, quotient groups. **DP**

Group Theory, P. *Representations of Solvable Groups.* Olaf Manz, Thomas R. Wolf. London Math. Soc. Lect. Note Ser., V. 185. Cambridge Univ Pr, 1993, xi + 302 pp, \$39.95 (P). [ISBN 0-521-39739-1]

Algebra, P. *Commutative Algebra: Syzygies, Multiplicities, and Birational Algebra.* Eds: William J. Heinzer, Craig L. Huneke, Judith D. Sally. Contemp. Math., V. 159. AMS, 1994, vii + 444 pp, \$61 (P). [ISBN 0-8218-5188-8] Proceedings of a 1992 AMS-IMS-SIAM Summer Research Conference.

Algebra, T(15–16: 1), C, L. *Learning Abstract Algebra with ISETL.* Ed Dubinsky, Uri Leron. Springer-Verlag, 1994, xix + 252 pp, \$49, with disk. [ISBN 0-387-94152-5] Consistent with the authors' view "that people learn best by *doing* and *thinking* about what they do," text treats basic algebraic structures, interwoven tightly with the use of ISETL (a mathematical programming language) to spur discovery and understanding. After introducing ISETL,

works through groups, subgroups, and the fundamental homomorphism theorem; then repeats process for rings. Concludes with factorization in integral domains, construction of splitting fields. Less formal depth than in some other currently popular texts. Stresses computer activities and exercises. JS

Algebra, P. *Discriminants, Resultants, and Multidimensional Determinants*. I.M. Gelfand, M.M. Kapranov, A.V. Zelevinsky. Math.: Theory & Applic. Birkhäuser, 1994, x + 523 pp, \$74.50. [ISBN 0-8176-3660-9] Studies discriminants and resultants of polynomials in several variables. Main approaches are geometric (projective duality and associated hypersurfaces), algebraic (homological algebras and determinants of complexes), combinatorial (Newton polytopes and triangulations). LC

Calculus, T(13: 1-3). *Calculus, Sixth Edition*. Earl W. Swokowski, et al. PWS, 1994, xxiv + 1384 pp. [ISBN 0-534-93624-5] New in this edition: more use of graphing technology; "Extended Problems and Group Projects" end chapters; biographical/historical sketches. (Fifth Edition, TR, May 1991.) AO

Calculus, T*(14: 1). *Basic Multivariable Calculus*. Jerrold E. Marsden, Anthony J. Tromba, Alan Weinstein. Springer-Verlag & WH Freeman, 1993, xv + 533 pp, \$39.95. [ISBN 3-540-97976-X; 0-7167-2443-X] Multivariable and vector calculus through theorems of Green, Gauss, and Stokes. Emphasizes intuitive understanding, computational skill. Careful exposition; many physical examples. AO

Real Analysis, T(17-18: 1), S, P, L. *A Concise Introduction to the Theory of Integration, Second Edition*. Daniel W. Stroock. Birkhäuser, 1994, viii + 184 pp, \$24.50. [ISBN 0-8176-3759-1] Treats both modern integration theory and advanced calculus. To blend the two, much space is devoted to \mathbb{R}^n -theory. New to this edition: two new sections on aspects of Lebesgue theory; minor reordering of material; solutions to some exercises. KS

Real Analysis, T(16-18), P, L. *The Riemann Approach to Integration: Local Geometric Theory*. Washek F. Pfeffer. Tracts in Math., V. 109. Cambridge Univ Pr, 1993, xv + 302 pp, \$49.95. [ISBN 0-521-44035-1] A detailed account of the McShane and Henstock-Kurzweil integrals (both generalized Riemann-Stieltjes integrals) and their relation. Classical results and recent developments: lipeomorphic change of variables, higher-dimensional multipliers, divergence theorem for discontinuously differentiable vector fields, etc. One-

dimensional theory treated first. Worthwhile exercises are integrated into exposition. KS

Partial Differential Equations, S, P. *PLTMG: A Software Package for Solving Elliptic Partial Differential Equations: Users' Guide 7.0*. Randolph E. Bank. Front. in Appl. Math., V. 15. SIAM, 1994, xii + 128 pp, \$24.50 (P). [ISBN 0-89871-330-7] Reference manual for PLTMG, a piecewise-linear finite element multigrid package, freely available from Netlib. AO

Partial Differential Equations, P. *Nonlinear Nonlocal Equations in the Theory of Waves*. P.I. Naumkin, I.A. Shishmarev. Transl. of Math. Mono., V. 133. AMS, 1994, ix + 289 pp, \$149. [ISBN 0-8218-4573-X]

Partial Differential Equations, T(18: 1), S, P. *Microlocal Analysis for Differential Operators: An Introduction*. Alain Grigis, Johannes Sjöstrand. London Math. Soc. Lect. Note Ser., V. 196. Cambridge Univ Pr, 1994, 151 pp, \$29.95 (P). [ISBN 0-521-44986-3] Assumes basics of distribution theory, functional analysis, differential geometry. Topics include pseudodifferential operators, local symplectic geometry, global theory of Fourier integral operators, spectral theory, Cauchy problems, wavefront sets, propagation of singularities. Concise treatment; chapter exercises. HD

Partial Differential Equations, T(17-18: 2). *Partial Differential Equations in Classical Mathematical Physics*. Isaak Rubinstein, Lev Rubinstein. Cambridge Univ Pr, 1993, xiv + 676 pp, \$94.95. [ISBN 0-521-42058-4] Theory of PDE's, viewed as language for describing continuous phenomena in physics. Basic natural science concepts are a thorough exposition of standard scientific PDE problems. A rigorous, systematic treatment of mathematics applied in classical physics. DS

Dynamical Systems, T(15-17: 2, 3), S, P, L. *Chaos and Nonlinear Dynamics: An Introduction for Scientists and Engineers*. Robert C. Hilborn. Oxford Univ Pr, 1994, xvii + 654 pp, \$55. [ISBN 0-19-505760-0] An introduction for scientists and engineers, by a physicist. Surveys and explains what nonlinear dynamics is about and what it does; not fully rigorous. Assumes introductory physics, calculus through differential equations. Analytic and computer exercises end each chapter. KS

Dynamical Systems, T(17: 1, 2), P, L. *Chaos in Dynamical Systems*. Edward Ott. Cambridge Univ Pr, 1993, xii + 385 pp, \$69.95. [ISBN 0-521-43215-4] For science and engineering students. Besides standard material, treats quasiperiodicity, Hamiltonian systems, multifractals, quantum chaos. Retains math-

ematical flavor despite relegating mathematical details to appendices. SK

Numerical Analysis, T(16-17: 1), L*. *Numerical Solution of Ordinary Differential Equations.* Lawrence F. Shampine. Chapman & Hall, 1994, x + 484 pp, \$64.95. [ISBN 0-412-05151-6] Theory and practice of numerical solution of initial value problems for ODE systems. Practical focus. Assumes background in differential equations, numerical analysis. AO

Analysis, S(18), P. *The Banach-Tarski Paradox.* Stan Wagon. Cambridge Univ Pr, 1993, xviii + 253 pp, \$24.95 (P). [ISBN 0-521-45704-1] The paradoxes at issue range from surprising (perfect squares and positive integers are equally numerous) to astonishing (any bounded set in space can be dissected and rearranged to form any other bounded set). Book readably explores such paradoxes and links to other areas, especially measure theory. Preface and brief addendum mention recent work, e.g., of Laczkovich on circle-squaring by translation. (1985 hardcover text, TR, March 1986.) PZ

Analysis, T(17-18: 1, 2), L. *Difference Equations: Theory and Applications, Second Edition.* Ronald E. Mickens. Van Nostrand Reinhold, 1990, xi + 448 pp, \$54.95. [ISBN 0-442-00136-3] New edition adds exercises and a new chapter on applications. (First Edition, TR, February 1989.) AO

Algebraic Geometry, S(18), P, L. *Theory of Algebraic Invariants.* David Hilbert. Cambridge Univ Pr, 1993, xiv + 191 pp, \$39.95; \$19.95 (P). [ISBN 0-521-44457-8; 0-521-44903-0] Translated and edited notes from a course of 51 lectures given by Hilbert in 1897, focused largely on his own fundamental work in invariant theory. A skillful, readable, coherent presentation, with a contemporary flavor. Main results include Hilbert's Basis Theorem, a Nullstellensatz, and the Syzygy Theorem. JS

Differential Geometry, P. *Harmonic Maps and Integrable Systems.* Eds: Allan P. Fordy, John C. Wood. Aspects of Math., V. E23. Friedr Vieweg & Sohn, 1994, 329 pp, \$64. [ISBN 3-528-06554-0] 12 expository articles explain applications of integrable systems to finding harmonic maps and related problems.

Geometry, T*(16-17: 1, 2), L*. *Visual Geometry and Topology.* Anatolij Fomenko. Springer-Verlag, 1994, xvi + 324 pp, \$89. [ISBN 0-387-53361-3] A beautiful textbook. Aims to "narrate, in an accessible and fairly visual language, some classical and modern achievements of geometry in both intrinsic mathematical problems and applications." Contains 287 illustrative drawings, plus 50 plates of

the author's original artwork. Topics include polyhedra, simplicial complexes, homologies, low-dimensional manifolds, symplectic topology, Hamiltonian mechanics, minimal surfaces, fractal geometry, and hyperbolic geometry. DP

Geometry, T*(14: 1), P, L. *College Geometry: A Discovery Approach.* David C. Kay. Harper-Collins, 1994, xvi + 494 pp, \$38. [ISBN 0-06-500006-4] Guided self-discovery of several Euclidean gems precedes axiomatic treatment of foundations and transformation approach. Concludes with hyperbolic geometry. Includes a wealth of figures, historical notes, graded exercises (including research projects); useful ancillary materials are available. An appealing text for mathematics majors and for prospective secondary teachers. JNC

Algebraic Topology, T(15-16: 1), L*. *A Geometric Introduction to Topology.* C.T.C. Wall. Dover, 1993, vi + 168 pp, \$6.95 (P). [ISBN 0-486-67850-4] A minimal approach, assumes no general topology and avoids simplices. Parts I and II culminate in a proof of Alexander's duality theorem. Part III illustrates connections with other branches of pure mathematics. (1972 Addison-Wesley edition, TR, August-September 1972; Extended Review, December 1975.) DP

Optimization, T(16-17: 1, 2), L. *Linear Programming.* James P. Ignizio, Tom M. Cavalier. Ser. in Industrial & Systems Eng. Prentice Hall, 1994, xx + 666 pp. [ISBN 0-13-183757-5] Introduction to formulation, solution, and analysis of linear programming models. Also covers network and integer models, multiobjective optimization. Relatively strong emphasis on fundamental concepts and mathematical theory; assumes some linear algebra. AO

Probability, P. *Conditional Measures and Applications.* M.M. Rao. Pure & Appl. Math., V. 177. Marcel Dekker, 1993, xiv + 417 pp, \$135. [ISBN 0-8247-8884-2] Axiomatic development of conditional measures from several different directions. Applications include computational problems associated with conditioning. Uses historical development as motivation. No exercises; extensive bibliography. MK

Probability, P. *Counterexamples in Probability and Real Analysis.* Gary L. Wise, Eric B. Hall. Oxford Univ Pr, 1993, xii + 211 pp, \$39.95. [ISBN 0-19-507068-2] Counter-intuitive examples from real analysis and measure theoretic probability. RSK

Stochastic Processes, P. *Simulation and Chaotic Behavior of α -stable Stochastic Processes.* Aleksander Janicki, Aleksander Weron. Pure & Appl. Math., V. 178. Marcel Dekker,

1994, vii + 355 pp, \$125. [ISBN 0-8247-8882-6]

Stochastic Processes, T(18: 1, 2), S, C. *Numerical Solution of SDE Through Computer Experiments.* Peter E. Kloeden, Eckhard Platen, Henri Schurz. Universitext. Springer-Verlag, 1994, xiv + 292 pp, \$49 (P), with disk. [ISBN 0-387-57074-8] Stochastic differential equations (SDEs) and discrete time approximation (strong and weak) of their solutions. Includes Turbo Pascal listings and disk. RWJ

Elementary Statistics, T(13-14: 1). *A First Course in Probability Models and Statistical Inference.* J.H.C. Creighton. Texts in Stat. Springer-Verlag, 1994, xxxi + 717 pp, \$49.95. [ISBN 0-387-94114-2] Probability, descriptive statistics, discrete and continuous probability models, estimation, hypothesis testing, and simple linear regression. Over 350 pages on hints and solutions of exercises. RWJ

Statistical Methods, T(15-17: 1), P. *Modeling Experimental and Observational Data.* Clifford E. Lunneborg. Duxbury Pr, 1994, xv + 544 pp. [ISBN 0-534-21426-6] For students in biological and behavioral sciences; emphasizes generality of the modeling approach in handling analysis of variance and regression problems, including logistic and Poisson regression. Techniques and diagnostic procedures are illustrated by meaningful examples. RSK

Statistical Methods, T(15-17: 1), S. *Visualizing Data.* William S. Cleveland. Hobart Pr, 1993, 360 pp, \$40. [ISBN 0-9634884-0-6] From Preface: "The success of a visualization tool should be based solely on the amount we learn about the phenomenon under study." By the same criteria, this text is a success. Through about 20 analyses (classic studies and recent data sets) text exposes various graphical methods and visualization tools, always keeping scientific goals in sight. Some standard techniques ($Q-Q$ plots, boxplots, parametric and non-parametric fits, slicing, 3-D plots, etc.), some less standard (stereograms, loess curve fitting, factor-plane methods, multi-way dot plots). Readable text, thoughtful examples, useful techniques. No exercises, but still a possible text for scientific data analysis. MK

Statistical Methods, T(18), S, P. *Statistical Models Based on Counting Processes.* Per Kragh Andersen, et al. Ser. in Stat. Springer-Verlag, 1993, xi + 767 pp, \$69. [ISBN 0-387-97872-0] Theoretical framework (e.g., counting processes, continuous-time martingales, stochastic integration) for event history/survival analysis. Covers applications and theory underlying many models and techniques. High math-

ematical level, but many examples and applications. Amazing variety of topics: from non-parametric estimation to frailty models. MK

Statistical Methods, T(16-17), C. *Statistical Principles of Research Design and Analysis.* Robert O. Kuehl. Duxbury Pr, 1994, xvi + 686 pp, with disk. [ISBN 0-534-18804-4] Classical introduction to design of experiments, stressing *research design*: "the total effort in a study that includes development of the research hypothesis, the choice of treatment design to address the research hypothesis, and the experiment design choice to facilitate efficient data collection." A beautifully written, well-conceived textbook for practitioners and newcomers. Examples based on actual studies from life sciences, agriculture, industrial and chemical engineering. Good exercises; includes disk. MK

Statistical Methods, P. *Model-Free Curve Estimation.* Michael E. Tarter, Michael D. Lock. Mono. on Stat. & Appl. Prob., V. 56. Chapman & Hall, 1993, x + 290 pp, \$49.95. [ISBN 0-412-04251-7-5]

Statistics, P, L. *The Collected Works of John W. Tukey, Volume VIII: Multiple Comparisons: 1948-1983.* Eds: Henry I. Braun, et al. Chapman & Hall, 1994, lxi + 485 pp, \$62.95. [ISBN 0-412-05121-4]

Statistics, T(17-18: 2); L. *Advanced Linear Models: Theory and Applications.* Song-Gui Wang, Shein-Chung Chow. Stat.: Textbooks & Mono., V. 141. Marcel Dekker, 1994, x + 537 pp, \$165. [ISBN 0-8247-9169-X] Statistical inference for linear models and unified coverage of regression, ANOVA, ANCOVA, and variance components models. Preliminary chapters treat matrix results and the multivariate normal and related distributions. RWJ

Statistics, T(17-18: 1), L. *Hilbert Space Methods in Probability and Statistical Inference.* Christopher G. Small, D.L. McLeish. Wiley, 1994, xi + 252 pp, \$59.95. [ISBN 0-471-59281-1] A Hilbert space approach to random variables. Applications to martingales, stochastic integration, interpolation, density estimation. Assumes linear algebra, probability, and statistics, and some knowledge of basic stochastic processes. RWJ

Programming, P, C. *The Borland C++ 4.0 Primer.* Keith Weiskamp. Academic Pr, 1994, xviii + 568 pp, \$39.95 (P), with disk. [ISBN 0-12-742683-3]

Programming, C*. *Mastering Mathematica: Programming Methods and Applications.* John Gray. Academic Pr, 1994, xx + 644 pp, (P), with disk. [ISBN 0-12-296040-8] Core subject is

different types of Mathematica programming. Includes introductory tutorials, chapters on representing mathematics in Mathematica, translating mathematical problems to Mathematica-solvable form. A must for prospective serious Mathematica users. DS

Computer Systems, T(16), S*, C, L*. *Applied Mathematica: Getting Started, Getting It Done.* William T. Shaw, Jason Tigg. Addison-Wesley, 1994, xiv + 432 pp, (P). [ISBN 0-201-54217-X] Impressive introduction to Mathematica's power and versatility. Useful to both neophytes (though not suitable as a first pass through Mathematica) and experienced users. Topics include data visualization in two and three dimensions, statistical analysis, time series, digital imaging. Includes basic Mathematica techniques, detailed coverage of data handling. No exercises, but many well thought-out examples. Highly recommended. MPR

Computer Systems, P. *Motif Programming Manual: Volume Six A.* Dan Heller, Paula M. Ferguson. O'Reilly & Assoc, 1994, xlii + 972 pp, \$34.95 (P). [ISBN 1-56592-016-3]

Applications (Behavioral Science), S(15-16), P. *Theory of Moves.* Steven J. Brams. Cambridge Univ Pr, 1994, xii + 248 pp, \$17.95 (P). [ISBN 0-521-45867-6; 0-521-45226-0] Described as a theory that is based upon and extends the classical theory of games, the work described in this book bears little resemblance to the kind of reasoning and proof mathematicians expect to see in a book on game theory. The effort to build a dynamic theory that anticipates the next move, and the next, makes interesting reading, as do examples that range from the Old Testament to the attempted coup against Gorbachev, but it won't do for understanding game theory. AWR

Applications (Biological Science), P. *Fractals: A User's Guide for the Natural Sciences.* Harold M. Hastings, George Sugihara. Oxford Univ Pr, 1993, xii + 235 pp, \$19.95 (P). [ISBN 0-19-854597-5] Applications of fractal geometry to the natural world. Treats mathematics of fractals and modeling patterns; includes case studies and user-ready programs. DH

Applications (Biological Science), P. *Cell Biology.* Eds: Byron Goldstein, Carla Wofsy. Lect. on Math. in the Life Sci., V. 24. AMS, 1994, x + 135 pp, \$38 (P). [ISBN 0-8218-1175-4] Invited lectures from the 1992 Symposium on Some Mathematical Questions in Biology held at the annual meeting of the American Society for Cell Biology.

Applications (Fluid Dynamics), P. *The Couette-Taylor Problem.* Pascal Chossat, Gérard Iooss. Appl. Math. Sci., V. 102.

Springer-Verlag, 1994, ix + 233 pp, \$44.50. [ISBN 0-387-94154-1] Local analysis of vortex flow of fluid between two concentric cylinders. Provides experimental results and describes G. Taylor's model to study these experiments. Emphasizes bifurcation and symmetry breaking arguments. DS

Applications (Physics), S(16-18), P, L. *The Special Theory of Relativity: A Mathematical Exposition.* Anadijiban Das. Universitext. Springer-Verlag, 1993, xii + 214 pp, \$39.95 (P). [ISBN 0-387-94042-1] Strong mathematical emphasis—physical situations are seldom mentioned. A useful reference, but a difficult entrance into the subject. MU

Applications (Physics), T*(15-16: 1), S, L. *A Brief on Tensor Analysis, Second Edition.* James G. Simmonds. Undergrad. Texts in Math. Springer-Verlag, 1994, xiv + 112 pp, \$29.95. [ISBN 0-387-94088-X] New edition corrects typos, augments exercises; new section treats differential geometry. A real gem. (*First Edition*, TR, February 1983.) MU

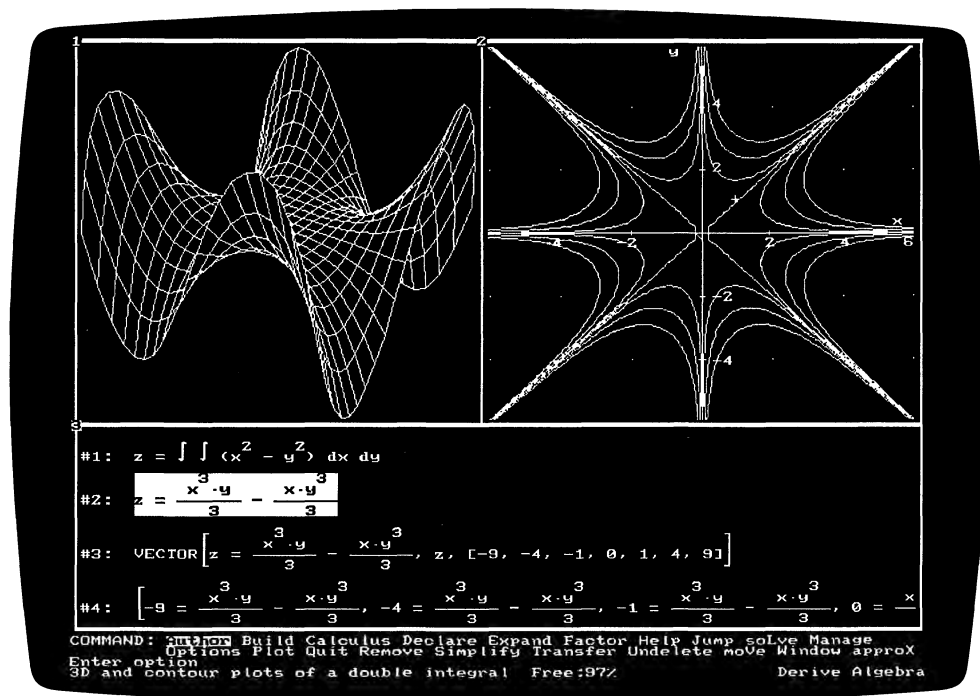
Applications, P, L*. *Bridging Mind and Model: Papers in Applied Mathematics.* Ed: Peter J. Costa. St. Thomas Tech Pr (Center for Appl. Math., Mail #4410, 2115 Summit Ave., St. Paul, MN 55105-1096), vi + 306 pp, \$24.95. [ISBN 0-9624229-7-5] 8 expository papers on maritime traffic control, path reconstruction, biochemistry, textile manufacturing, laser physics, and signal processing.

Applications, T(17-18: 1), P. *Mathematical Foundations of Elasticity.* Jerrold E. Marsden, Thomas J.R. Hughes. Dover, 1993, xviii + 556 pp, \$14.95 (P). [ISBN 0-486-67865-2] Mathematical foundations of 3-dimensional elasticity using modern differential geometry and functional analysis. For mathematicians, engineers, and physicists. Only advanced calculus is needed for chapters on kinematics and dynamics of continuous media; more background needed to study variational principles and functional analysis. (1983 Prentice-Hall text, TR, October 1983.) DS

Reviewers

JNC: Judith N. Cederberg, St. Olaf; LC: Laura Chihara, St. Olaf; HD: Hung Dinh, Macalester; SG: Steven Galovich, Carleton; TH: Tom Halverson, Macalester; DH: Deanna Haunsperger, Carleton; RWJ: Roger W. Johnson, Carleton; MK: Michael Kahn, St. Olaf; SK: Steve Kennedy, Carleton; RSK: Richard S. Kleber, St. Olaf; RM: Richard Molnar, Macalester; AO: Arnold Ostebee, St. Olaf; DP: David Peifer, St. Olaf; MPR: Matthew P. Richey, St. Olaf; AWR: A. Wayne Roberts, Macalester; KS: Karen Saxe, Macalester; JS: John Schue, Macalester; DS: Dan Schwalbe, Macalester; MU: Milton Ulmer, Carleton; MW: Martha Wallace, St. Olaf; PZ: Paul Zorn, St. Olaf.

NEW ~~DERIVE~~ VERSION 3!



DERIVE is a powerful computer algebra system for doing symbolic and numeric mathematics on your personal computer.

DERIVE:

- Performs numerical operations exactly with no round-off error
- Approximates irrational expressions to thousands of digits of precision
- Algebraically simplifies, expands, and factors expressions; and solves equations
- Applies the rules of trigonometry, calculus,

matrix algebra, and vector calculus

- Plots explicitly and implicitly defined functions in 2D with zooming and auto-scaling
- Generates 3D wire-frame function plots using hidden-line removal
- Displays and prints expressions using standard 2D mathematical notation
- Provides an easy to use, menu-driven interface with on-line help
- Is ideal for students, teachers, engineers, scientists, and mathematicians

DERIVE Requirements

(regular memory version):

A PC compatible running MS-DOS with 512K memory and a 3 1/2 inch (720K) diskette drive *or* an HP 95LX, 100LX or 200LX palmtop computer with 1M memory and connectivity pack for downloading. List \$125.

DERIVE XM Requirements

(extended memory version):

A 386, 486 or Pentium® based PC compatible running MS-DOS version 3.0 or later with at least 2M of extended memory and a 3 1/2 inch (1.4M) diskette drive. List \$250.

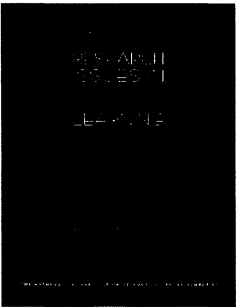
DERIVE is a registered trademark of Soft Warehouse, Inc.

 **Soft Warehouse**
HONOLULU • HAWAII

Soft Warehouse, Inc. • 3660 Waiialae Ave.
Ste. 304 • Honolulu, HI, USA 96816-3236
Ph: (808) 734-5801 • Fax: (808) 735-1105

Research Issues in Undergraduate Mathematics Learning Preliminary Analyses and Reports

James J. Kaput and Ed Dubinsky, Editors



Research in undergraduate mathematics education is important for all college and university mathematicians. If our students are to be more successful in understanding mathematics, then college faculty need to understand how mathematics is learned. This knowledge can guide us in curriculum reform and in improving our own teaching. It can help us make mathematics accessible to all students and it can increase the number of graduate students in mathematics.

This volume of research in undergraduate mathematics education informs us about the nature of student learning in some of the most important topics in the undergraduate curriculum: sets, functions, calculus, statistics, abstract algebra and problem solving. Paying careful attention to the trouble students have in learning mathematics will help us to work with students so they can deal with those difficulties.

A survey of the literature begins the volume. Becker and Pence have brought together an unusually complete list of references on research in collegiate mathematics. Their comments will guide those attempting to begin or to continue a program of research in student learning.

The sad fact that even good calculus students stumble over nonroutine problems is the theme of Selden, Selden, and Mason. Their conclusions point to significant shortcomings in the curriculum. This study of student difficulties is

continued by Ferrini-Mundy and Graham who investigate a single student's interactions with the fundamental concepts of the calculus. Baxter studies a group of students to learn how they acquire the concept of set, while Cuoco does the same for the concept of function.

Cooperative learning does help the student. That is the conclusion of Bonsangue, who investigates how two carefully matched classes of students in a statistics course perform on exams. How students learn to write proofs in group theory is the subject considered by Hart. Rosamond breaks new ground by comparing how emotions vary in their effect on the problem solving ability of novices and experts.

All college faculty should read this book to find how they can help their students learn mathematics.

150 pp., Paperbound, 1994

ISBN 0-88385-090-7

List: \$24.00

Catalog Number NTE-33

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Membership Code _____

Name _____

Address _____

City _____

State _____ Zip Code _____

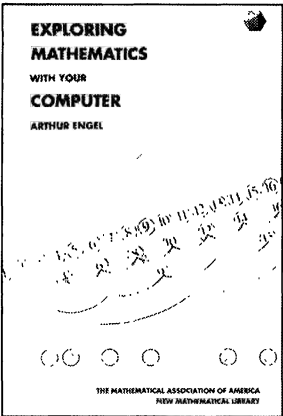
Qty.	Catalog Number	Price
Total \$		
Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
Credit Card No. _____		
Signature _____		
Exp. Date _____		

Exploring Mathematics With Your Computer

Arthur Engel

A must for academic libraries supporting an undergraduate major in mathematics. Public libraries with strong science collections should have this book as a resource for traditional mathematics topics and as a recreation for the mathematically inclined. Capable high school students would also benefit.

—CHOICE



Today's personal computer gives its owner tremendous power which can be used for experimental investigations and simulations of unprecedented scope, leading to mini-research. This book is a first step into this exciting field.

This is a mathematics book, not a programming book, although it explains Pascal to beginners. It is aimed at high school students and undergraduates with a strong interest in mathematics and teachers looking for fresh ideas. It is full of diverse mathematical ideas requiring little background. It includes a large number of challenging problems, many of which illustrate how numerical computation leads to conjectures which can then be proved by mathematical reasoning.

You will find 65 interesting and substantial mathematical topics in this book, and over 360 problems. Each topic is illustrated with examples and corresponding programs. The major goal of the book is to use the computer to collect data and formulate conjectures suggested by the data.

It is assumed that readers have a PC at their disposal.

264 pp., Paperbound, 1993

ISBN 0-88385-636-0

List: \$38.00 MAA Member: \$26.50

Catalog Number NML-35

A 3.5" IBM-compatible disk containing the Pascal programs described in the book is packaged with this volume.

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

Foreign Orders Please add \$3.00 per item ordered to cover postage and handling fees. The order will be sent via surface mail. If you want your order sent by air, we will be happy to send you a proforma invoice for your order.

Membership Code	Qty.	Catalog Number	Price

Name _____			
Address _____			
City _____			
State _____ Zip Code _____			
			Total \$ _____
			Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
			Credit Card No. _____
			Signature _____ Exp. Date _____

Student Research Projects in Calculus

Marcus Cohen, Edward D. Gaughan, Arthur Knoebel, Douglas S. Kurtz, and David Pengelley

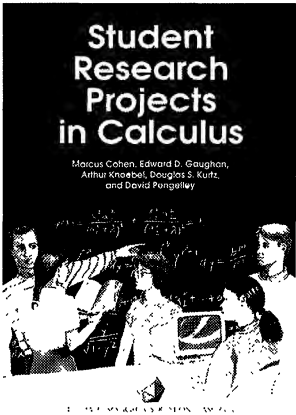
It is yet another workable way to rejuvenate college calculus. The monograph is highly recommended to those interested in experimenting with mathematics curricula. —AAAS, Science Books & Films

I found the book very readable and thought provoking. Whether your students are in engineering, pure or applied science, or even liberal arts, this book may change the way you teach—and the way they learn—calculus for the better! —The Mathematics Teacher

An important contribution to the growing list of new curricular materials becoming available to assist in the teaching of calculus. —CHOICE

Changing the way students learn calculus was the goal of the authors of this excellent guidebook. In the Spring of 1988, they began work on a student project approach to calculus.

You can use their methods in teaching your own calculus courses. Over 100 projects are presented, all of them ready to assign to your students in single and multivariable calculus. The projects were designed with one goal in mind: to get students to think for themselves. Each project is a multistep, take home problem, allowing students to work both individually and in groups.



Each project has accompanying notes to the instructor reporting students' experiences. The notes contain information on prerequisites, list the main topics the project explores, and suggest helpful hints. The authors have also provided several introductory chapters to help instructors use projects successfully in their classes and begin to create their own.

232 pp., Paperbound, 1992

ISBN 0-83385-503-8

List: \$25.50 MAA Member: \$18.00

Catalog Number SRPC

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

-----	Qty.	Catalog Number	Price
Membership Code	-----		
-----	-----		
Name _____	Total \$ _____		
Address _____	Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
City _____	Credit Card No. _____		
State ____ Zip Code _____	Signature _____		
	Exp. Date _____		

*Revised and expanded
second edition*

Polyominoes

Puzzles, Patterns, Problems, and Packings

Solomon W. Golomb

Cloth: \$24.95 ISBN 0-691-08573-0

The Enjoyment of Math

**Hans Rademacher
and Otto Toeplitz**

Paper: \$12.95 ISBN 0-691-02351-4

New paperback edition

The Mathematical Career of Pierre de Fermat, 1601-1665

Michael Sean Mahoney

Paper: \$18.95 ISBN 0-691-03666-7

e: The Story of a Number

Eli Maor

Cloth: \$24.95 ISBN 0-691-03390-0

Ramification Theoretic Methods in Algebraic Geometry

Shreeram Abhyankar

Paper: \$25.00 ISBN 0-691-08023-2

Temperley-Lieb Recoupling Theory and Invariants of 3-Manifolds

**Louis H. Kauffman and
Sostenes Lins**

Paper: \$22.50 ISBN 0-691-03640-3

Cloth: \$49.50 ISBN 0-691-03641-1

An Introduction to G-Functions

**Bernard Dwork,
Giovanni Gerotto,
and Francis J. Sullivan**

Paper: \$29.95 ISBN 0-691-03681-0

Cloth: \$59.50 ISBN 0-691-03675-6

Introduction to Ergodic Theory

Preliminary Informal Notes of University
Courses and Seminars in Mathematics

Ya. G. Sinai

Paper: \$29.95 ISBN 0-691-08182-4

Topics in Ergodic Theory

Ya. G. Sinai

Cloth: \$45.00 ISBN 0-691-03277-7

Essays on Fourier Analysis in Honor of Elias M. Stein

**Edited by Charles Fefferman,
Robert Fefferman,
and Stephen Wainger**

Cloth: \$65.00 ISBN 0-691-08655-9

Introduction to Arithmetic Theory of Automorphic Forms

Goro Shimura

\$39.50 ISBN 0-691-08092-5

PRINCETON UNIVERSITY PRESS

AVAILABLE AT FINE BOOKSTORES OR DIRECTLY FROM THE PUBLISHER: 609-883-1759

POLYOMINOES:

Puzzles and Problems in Tiling

George Martin

George Martin has done a truly marvelous job of presenting the material in this book in an attractive and clear way.

—Martin Gardner

It brings together results which are scattered throughout the literature and will be a standard reference for many years.

—Mathematical Reviews

Martin has an easy style of writing, and the material is effortless to read and understand.

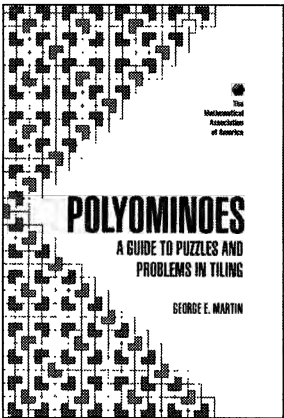
—Mathematics Teacher

The writing is clear, with enough formality to make the arguments make sense. This book is a necessary part of a well-stocked library in recreational mathematics.

—AAAS, Science Books & Films

Polyominoes will delight not only students and teachers of mathematics at all levels, but will be appreciated by anyone who likes a good geometric challenge. There are no prerequisites. If you like jigsaw puzzles or if you hate jigsaw puzzles but have ever wondered about the pattern of some floor tiling, there is much here to interest you.

A polyomino is a shape cut along the lines from square graph paper; the pronunciation of *polyonimo* begins as does *polygon* and ends as does *domino*. Tilings, also called tessellations of mosaic patterns, are older than civilization itself. Tiling with polyominoes provides challenges that range from the popular jigsawlike puzzles to easily



understood mathematical research problems. You will find unsolved puzzles and problems of both kinds here. Answers are provided for most of the problems that have a known solution.

It is only fair to repeat here the warning stated in the preface to this book, "Playing with polyominoes can be habit forming."

172 pp., Paperbound, 1991
ISBN 0-88385-501-1
List: \$22.00 MAA Member: \$17.00
Catalog Number: POLY

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-800-331-1622 Fax (202) 265-2384

Foreign Orders Please add \$3.00 per item ordered to cover postage and handling fees. The order will be sent via surface mail. If you want your order sent by air, we will be happy to send you a proforma invoice for your order.

Membership Code	_____
Name	_____
Address	_____
City	_____
State	_____
Zip Code	_____

Qty.	Catalog Number	Price
_____	_____	_____
Total \$		_____
Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
Credit Card No. _____		
Signature _____		Exp. Date _____

Symbolic Computation in Undergraduate Mathematics Education

Zaven Karian, Editor

If you are considering putting a symbolic computing system into your curriculum, this is one publication you should have.

—Mathematics Teacher

This well-written book should be helpful to anyone using symbolic computation as an aid in teaching undergraduates—The book provides a number of examples for presenting probability and statistics in a way that removes the tedium and emphasizes the underlying ideas.

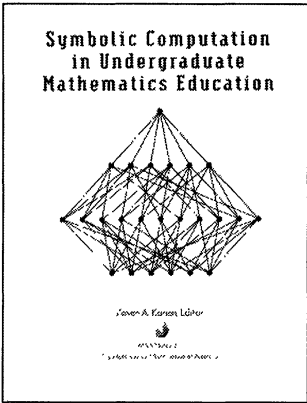
—AAAS, Science Book and Films

If you have any plans to integrate symbolic computing into your program, read and study this book first. Your students will thank you for it.

—AMATYC Review

This volume brings together many of the facets associated with the pedagogic uses of symbolic computation.

Part I consists of articles that deal with general issues of learning mathematics and the role of symbolic computation in that process. The articles in Part II describe the use of symbolic computation in teaching calculus. Some of the areas covered are the use of symbolic computation in a laboratory calculus course, the uses of Derive in the instruction of calculus, antidifferentiation and the



definite integral, and the experiences and reflections of teachers who have used symbolic computation in calculus instruction.

Part III consists of papers on sophomore-level courses on linear algebra and differential equations. The articles in Part IV describe what can be done in using symbolic computation in teaching combinatorics, probability and statistics courses. The articles and references in Part V will help you get started in using some of these ideas at your own institution.

200 pp., 1992, Paperbound

ISBN 0-88385-082-6

List: \$24.00

Catalog Number NTE-24

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 2003
1-800-331-1622 Fax (202) 265-2384

Membership Code _____		Qty.	Catalog Number	Price
Name _____		_____		
Address _____		_____		
City _____		Total \$ _____		
State _____ Zip Code _____		Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD		
		Credit Card No. _____		
		Signature _____ Exp. Date _____		

FROM MARTIN GARDNER

Mathematical Magic Show

... We visit most of the prime sites of recreational mathematics: game theory, factorials, puzzles, playing cards, finger arithmetic, Möbius bands, polyominoes, perfect numbers, the knight's tour, trees, and dice. Gardner always has new facts and ideas to add interest to even the most well-trodden areas. —Times Literary Supplement

312 pp., Paperbound, 1990
ISBN 0-88385-449-X
List: \$19.50 MAA Member: \$16.50
Catalog Number MAGIC

Mathematical Carnival

His startling gift for bringing the sublime to the people is unabated. Once again, hard mathematical ideas are conveyed with fluency, charm, and utter clarity. As a philosopher, I warmly salute his judgement of when to leave a metaphysical question enticingly open. His craftsmanship remains exquisite. —New Scientist

320 pp., Paperbound, 1988
ISBN 0-88385-448-1
List: \$18.00 MAA Member: \$15.00
Catalog Number MCR

Riddles of the Sphinx and Other Mathematical Puzzle Tales

This book charms, informs, inspires, puzzles, and delights, and the reader can dip in almost anywhere and get hooked by the natural lucidity of style and the friendly tone which are so characteristic of Martin Gardner.

—Mathematical Spectrum

184 pp., Paperbound, 1987
ISBN 0-88385-632-8
List: \$16.00 MAA Member: \$13.00
Catalog Number NML-32

Mathematical Circus

A circus suggests fun and enjoyment and there is plenty of both to be found here. The book should certainly be in the school library. It will also be valuable resource for the teacher.

—The Mathematical Gazette

300 pp., Paperbound, 1992
ISBN 0-88385-506-2
List: \$19.50 MAA Member: \$16.50
Catalog Number CIRCUS

ORDER FROM:

The Mathematical Association of America
1529 Eighteenth Street, NW
Washington, DC 20036
1-(800) 331-1622 Fax (202) 265-2384

Membership Code	Qty.	Catalog Number	Price

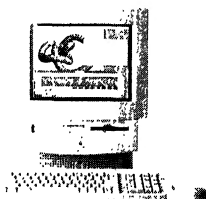
Name _____			
Address _____			
City _____			Total \$ _____
State _____ Zip Code _____			Payment <input type="checkbox"/> Check <input type="checkbox"/> VISA <input type="checkbox"/> MASTERCARD
			Credit Card No. _____
			Signature _____ Exp. Date _____

PUT ONE OF THESE

Customize a *Mathematica*
Site License for Your Campus
1-800-441-MATH (6284)



ON EVERY ONE OF THESE



FOR JUST ONE OF THESE.



How are universities around the world putting *Mathematica*® on every computer on campus for

as little as \$1 per student? They are taking advantage of new *Mathematica* site

license programs. In fact, site licenses at over 700 universities have made *Mathematica* accessible to millions of students without breaking the school budget.

This new series of flexible, affordable site license programs

puts you in charge. You choose where you want *Mathematica*, what kinds of computers you want it on, how you want to network it in your labs, and how much you want to invest. And you can save up to 90% off the already-reduced academic price.

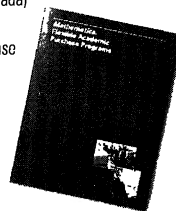
If you're excited about preparing more students for tomorrow by teaching with the world's leading technical computing system today, talk to us about a *Mathe-*

matica site license. We'll work together to customize one to fit your school's needs.

1-800-441-MATH (6284)

(U.S. and Canada)

Call us about a site license for your campus and ask for a free copy of this booklet, *Mathematica Flexible Academic Purchase Programs*.



Wolfram Research

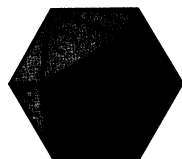
Mathematica is available for: Macintosh • Power Macintosh • Microsoft Windows • Microsoft Windows NT • MS-DOS • Sun SPARC • HP • Hitachi • DEC Alpha OSF/1, RISC, VAX/VMS • IBM RISC • SGI • NEC PC • NEC EWS • NEXTSTEP • CONVEX • and others.

Corporate headquarters: **Wolfram Research, Inc.**, +1-217-398-0700, fax: +1-217-398-0747, email: info@wri.com. Europe: **Wolfram Research Europe Ltd.**, +44-(0)1993-883400; fax: +44-(0)1993-883800; email: info-euro@wri.com. Asia: **Wolfram Research Asia Ltd.** (Tokyo office), +81-(0)3-5276-0506; fax: +81-(0)3-5276-0509; email: info-asia@wri.com.

© 1994 Wolfram Research, Inc. *Mathematica* is a registered trademark of Wolfram Research, Inc. *Mathematica* is not associated with Mathematica Policy Research, Inc. or MathTech, Inc. All other product names mentioned are trademarks of their producers.

The American Mathematical Monthly

Volume 101, Number 10 / DECEMBER 1994
(ISSN 0002-9890)



Contents

ARTICLES

- The Rectilinear Crossing Number of a Complete Graph and Sylvester's "Four Point Problem" of Geometric Probability / EDWARD R. SCHEINERMAN and HERBERT S. WILF 939
- String Matching for the Novice / ORA E. PERCUS and JEROME K. PERCUS 944
- Bernoulli Trials and Number Theory / DON RAWLINGS 948
- What Is the Shape of a Mylar Balloon? / WILLIAM H. PAULSEN 953
- Euler's Theorem for Polyhedra: A Topologist and Geometer Respond / PETER HILTON and JEAN PEDERSEN 959
- The Role of Paradoxes in the Evolution of Mathematics / I. KLEINER and N. MOVSHOVITZ-HADAR 963
- Regions in the Complex Plane Containing the Eigenvalues of a Matrix / RICHARD A. BRUALDI and STEPHEN MELLENDORF 975
- Characterization of Solvable Quintics $x^5 + ax + b$ / BLAIR K. SPEARMAN and KENNETH S. WILLIAMS 986
- A Halmos Problem and a Related Problem / JOHN B. COSGRAVE 993
-

FEATURES

COMMENTS 938

NOTES

- A "Popular" Class Number Formula / KURT GIRSTMAIR 997
- Variations on Wolstenholme's Theorem / EMRE ALKAN 1001
- The Second-Partials Test for Local Extrema of $f(x, y)$ / LEONARD GILLMAN 1004

UNSOLVED PROBLEMS

- Mousetrap / RICHARD K. GUY and RICHARD J. NOWAKOWSKI 1007

THE AUTHORS 1011

PROBLEMS AND SOLUTIONS 1013

REVIEWS

- Linear Programs and Related Problems*, by Evar D. Nering and Albert W. Tucker / STEPHEN B. MAURER 1022

TELEGRAPHIC REVIEWS 1027

INDEX TO AMERICAN MATHEMATICAL MONTHLY VOLUME 101 1033

